Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

An Explainable Attention-Gated CNN–Transformer Framework for Automated Microplastic Recognition in Freshwater Microscopy Images

Majid Hajibarati

hajibaratimajid@gmail.com

Islamic Azad University

Research Article

Keywords: Environmental monitoring, Particle morphology, Deep learning, Vision Transformer, Plastic pollution, Machine learning

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11105292/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

An Explainable Attention-Gated CNN–Transformer Framework for Automated Microplastic Recognition in Freshwater Microscopy Images

Majid Hajibarati

Faculty of Agriculture, Islamic Azad University, Tehran, Iran

Corresponding author: Majid Hajibarati
Email: hajibaratimajid@gmail.com
Phone: +989190769963
ORCID: 0009-0007-1971-1485

Original Research Article

Acknowledgments
Not applicable.

Abstract

Microplastic pollution requires rapid and reproducible monitoring methods that can complement labor-intensive visual and spectroscopic analyses. This study develops a morphology-aware artificial-intelligence framework for automated classification of field-collected microplastic particles and diatom algae in optical microscopy images. A public freshwater dataset containing 1,140 images from four classes (algae, filament, fragment, and pellet) was analyzed. After image-quality auditing, 1,137 unique images were retained for morphology–texture analysis, while deep-learning experiments used the original image collection under a fixed train/validation/test protocol. Handcrafted morphology and texture descriptors were evaluated with Random Forest, Extra Trees, radial-basis-function Support Vector Machine, and Histogram Gradient Boosting. Histogram Gradient Boosting achieved a cross-validation macro-F1 of 0.9341 ± 0.0255 and a held-out test accuracy of 0.9342. Deep-learning experiments compared ResNet50, Swin Transformer, and a CNN–Transformer fusion model combining ResNet50 and Swin representations. The fusion model achieved an accuracy of 0.9240, macro-F1 of 0.9222, and macro-AUC of 0.9954 on the 171-image test subset. Class-wise AUCs were 1.0000, 0.9950, 0.9886, and 0.9982 for algae, filament, fragment, and pellet, respectively. The results show that morphology-aware machine learning and complementary CNN–Transformer representations can support automated freshwater microplastic screening and provide a reproducible basis for environmental monitoring workflows.

Keywords: Environmental monitoring; Particle morphology; Deep learning; Vision Transformer; Plastic pollution; Machine learning

Introduction

Microplastics (MPs), generally defined as plastic particles smaller than 5 mm, are pervasive contaminants in marine, freshwater, terrestrial, atmospheric, food-related, and biological environments. Their persistence, heterogeneous origins, and interactions with organisms and other contaminants have made reliable identification and characterization essential for environmental monitoring and risk assessment. However, considerable heterogeneity remains in MP sampling, preparation, detection, classification, and reporting, limiting comparability among studies and complicating standardized monitoring (Goyetche et al., 2023; Yang et al., 2026; Scheiterlein and Fiener, 2026; Fernandez et al., 2026).

Conventional MP characterization combines optical or fluorescence microscopy with techniques such as scanning electron microscopy, Fourier-transform infrared (FTIR), Raman, and other spectroscopic approaches. Although these methods can provide detailed morphological or polymer-specific information, they may require expensive instrumentation, extensive sample preparation, specialized personnel, and substantial analysis time, and can be affected by background interference, weak signals, complex environmental matrices, and computationally intensive spectral imaging (Cowger et al., 2020; Goyetche et al., 2023; Höppener et al., 2023; Kim et al., 2024; Vitali et al., 2024; Yang et al., 2026; Smolen et al., 2025). Image-based analysis therefore provides an attractive complementary strategy because particle size, shape, elongation, circularity, surface irregularity, color, and texture can be quantified directly. Computer-vision and machine-learning approaches have enabled automated counting, morphology assessment, fluorescence-based particle detection, semantic segmentation, and rapid differentiation of fibers and fragments (Massarelli et al., 2021; Maes et al., 2017; Prata et al., 2020; Meyers et al., 2022; Vitali et al., 2024; Tsuchiya et al., 2025).

Deep learning has further expanded automated MP analysis by learning visual representations directly from microscopy images. Previous studies have investigated U-Net-derived segmentation, fluorescence-microscopy quantification, semantic segmentation of electron micrographs, weakly supervised segmentation, morphology extraction, Mask R-CNN-based categorization, feature-fusion architectures, and convolutional approaches for irregular environmental particles (Lee and Jhang, 2021; Park et al., 2022; Shi et al., 2022; Phan et al., 2023; Huang et al., 2023; Yao et al., 2025; Xu and Wang, 2024; Royer et al., 2024). Nevertheless, recognition remains sensitive to dataset quality, background structures, filter meshes, particle transparency, degradation state, and imaging conditions, indicating that strong performance on a controlled dataset does not necessarily ensure generalization across environmental matrices (Lee et al., 2024; Fernandez et al., 2026).

Particle morphology is also environmentally relevant because shape influences transport, settling, exposure, ingestion, and interactions with surrounding matrices. Quantitative morphology therefore provides information beyond categorical prediction. Previous studies have demonstrated the value of segmentation-derived physical descriptors, large-scale deep-learning shape classification, morphology-based categorization, automated shape recognition, and transfer learning for optical microscopy images (Phan et al., 2023; Liu et al., 2025; Huang et al., 2023; Shi et al., 2022; Harde and B, 2026). Other artificial-intelligence approaches have incorporated holographic imaging, polarization-sensitive imaging, portable digital holography, and polarized-light backscattering to obtain complementary structural information (Bianco et al., 2020; Li et al., 2024; Abimbola et al., 2025; Saur et al., 2026). Portable and real-time systems have also combined object detection and tracking, low-cost fluorescence imaging, spectroscopy, machine learning, similarity learning, and hyperspectral imaging for automated MP analysis (Sarker et al., 2024; Rermborirak et al., 2025; Kim et al., 2024; Höppener et al., 2023; Smolen et al., 2025; Yang et al., 2026).

Despite this progress, many existing systems focus on a single task, such as classification, segmentation, polymer identification, counting, or morphology measurement. Classical morphology descriptors are interpretable but may not capture complex spatial information, whereas convolutional neural networks effectively learn local texture and boundary patterns but may not fully represent broader spatial relationships. Transformer-based visual models provide a complementary mechanism for modeling longer-range dependencies, suggesting that CNN and Transformer representations may provide complementary information for heterogeneous microscopy images. This issue is particularly relevant to the field-collected freshwater dataset introduced by Almeida et al. (2025), comprising 1,140 optical microscopy images of algae, filaments, fragments, and pellets. Previous studies have demonstrated both the feasibility of deep-learning-based MP classification and the importance of robust preprocessing and morphology-aware analysis (Almeida et al., 2025; Lorenzo-Navarro et al., 2021; Giardino et al., 2023).

Accordingly, the present study develops an integrated morphology-aware artificial-intelligence framework for automated analysis of field-collected microplastics and diatom algae in optical microscopy images. The workflow combines data-quality auditing, quantitative morphology and texture characterization, interpretable machine learning, convolutional and Transformer-based deep representations, and cross-validation-based performance assessment. In addition to conventional classification metrics, class-specific errors, feature importance, prediction stability, and

model interpretability are evaluated. An attention-gated CNN–Transformer fusion strategy combining ResNet50 and Swin Transformer is investigated to determine whether adaptive integration of complementary local and global image representations can improve discrimination among algae, filaments, fragments, and pellets.

Materials and methods

Dataset and image preprocessing

A publicly available optical microscopy dataset of field-collected microplastics and diatom algae was used in this study. The samples originated from sediment collected from the Alagados freshwater reservoir in southern Brazil and represent environmentally collected particles with substantial variation in size, shape, transparency, color, and surface characteristics (Almeida et al., 2025). The original dataset comprised 1,140 RGB images distributed among algae (n = 300), filament (n = 300), fragment (n = 300), and pellet (n = 240). Diatom algae were retained as a separate non-plastic class because their biosilica structures may persist after sample preparation and visually resemble plastic particles during screening (Almeida et al., 2025).

Before the morphology–texture machine-learning analysis, all images were audited for readability, RGB consistency, class assignment, dimensions, and exact duplicate content. Three redundant images were identified and removed, resulting in 1,137 unique images for the morphology–texture analysis. The separately conducted deep-learning experiments used the original 1,140-image dataset. For deep-learning experiments, images were converted to RGB, resized to 224 × 224 pixels, and normalized using ImageNet statistics. Training-only augmentation included random rotation, translation, scaling, shear transformation, and horizontal flipping; no stochastic augmentation was applied to validation or test images. Comparable augmentation and transfer-learning strategies have been used in previous microplastic image-classification studies (Shi et al., 2022; Huang et al., 2023; Almeida et al., 2025; Liu et al., 2025; Harde and B, 2026).

For the morphology–texture machine-learning analysis, the duplicate-free dataset was stratified into an 80% development subset (n = 909) and a 20% independent held-out test subset (n = 228). Deep-learning experiments were conducted separately using a fixed 70%/15%/15% train/validation/test partition, producing a 171-image test subset. A fixed random seed of 42 was used for reproducibility. Class distributions for the morphology–texture machine-learning analysis are summarized in Table 1.

Table 1 Dataset composition and leakage-controlled data partitioning

Class	Original images	Unique images	Development set	Held-out test set
Algae	300	300	240	60
Filament	300	300	240	60
Fragment	300	299	239	60
Pellet	240	238	190	48
Total	**1,140**	**1,137**	**909**	**228**

Note. Exact duplicate images were removed before data partitioning. The duplicate-free dataset was stratified into an 80% development subset and a 20% independent held-out test subset.

Morphological and texture analysis

Particle morphology was quantified using an algorithmic pseudo-segmentation procedure because the source dataset provides image-level labels but no manually annotated pixel-wise masks. Similar segmentation-based approaches

have been used for quantitative microplastic morphology analysis and interpretable classification (Phan et al., 2023; Huang et al., 2023; Shi et al., 2022). The image background was estimated from border pixels in CIELAB color space, and foreground likelihood was derived from color difference and gradient information. Masks were refined by thresholding, morphological operations, connected-component filtering, and contour filling, after which the dominant particle region was selected. Because no pixel-level ground truth was available, Dice and intersection-over-union metrics were not calculated.

Geometric descriptors included area, perimeter, major- and minor-axis lengths, aspect ratio, eccentricity, circularity, extent, solidity, convexity, skeleton length, mean skeleton width, and boundary roughness. These descriptors provide interpretable information on particle shape and can support discrimination among microplastic morphologies (Phan et al., 2023). Aspect ratio is also environmentally relevant because particle geometry influences settling, transport, deposition, and environmental fate (Salehin et al., 2026).

Circularity was calculated as (Morét-Ferguson et al., 2010; Phan et al., 2023):

$$\text{Circularity} = (4 \times \pi \times A) / P^2 \quad (1)$$

where A is particle area and P is perimeter. Aspect ratio, extent, and solidity were calculated as:

$$\text{Aspect Ratio} = L_{major} / L_{minor} \quad (2)$$

$$\text{Extent} = A / A_{bbox} \quad (3)$$

$$\text{Solidity} = A / A_{convex} \quad (4)$$

where L_{major} and L_{minor} denote the major- and minor-axis lengths, A_{bbox} is the bounding-box area, and A_{convex} is the convex-hull area. These measures are consistent with quantitative morphology descriptors previously applied to interpretable microplastic shape analysis (Phan et al., 2023).

Texture and intensity characteristics included grayscale mean and standard deviation, entropy, edge density, Laplacian variance, RGB and HSV statistics, and gray-level co-occurrence matrix (GLCM) contrast, dissimilarity, homogeneity, energy, and correlation. Fractal-dimension and Hu-moment descriptors were additionally calculated to characterize boundary and shape complexity. Because a uniform pixel-to-micrometer calibration was unavailable, dimensional measurements were retained in pixel-based or relative units.

Interpretable machine learning and deep-learning models

Four conventional machine-learning algorithms were evaluated using the extracted morphology and texture descriptors: Random Forest, Extra Trees, radial-basis-function Support Vector Machine (RBF-SVM), and Histogram Gradient Boosting. Tree-based methods were selected for their ability to model nonlinear relationships and provide interpretable feature relevance, whereas RBF-SVM was included as a strong nonlinear classifier for relatively small tabular datasets. Similar feature-based approaches have demonstrated effective and interpretable discrimination of microplastic morphologies (Phan et al., 2023).

Feature standardization was applied before RBF-SVM training, while tree-based models used the original numerical feature scales. Class-balancing mechanisms were applied where supported. Models were compared using stratified five-fold cross-validation, and Random Forest feature importance was used to identify the morphology and texture variables most strongly associated with class discrimination.

Deep-learning experiments used ImageNet-pretrained transfer-learning architectures. ResNet50 served as the principal CNN baseline and the convolutional branch of the proposed fusion framework, while Swin Transformer was evaluated as the hierarchical attention-based model and Transformer branch. Previous microplastic studies have shown that architecture selection and transfer learning can substantially influence classification performance (Liu et al., 2025; Harde and B, 2026).

For each pretrained model, the original classification layer was replaced with a task-specific four-class head incorporating dropout and fully connected layers. All deep-learning architectures were trained under the same partitioning, augmentation, optimization, and evaluation framework to enable consistent comparison.

Proposed attention-gated CNN–Transformer fusion model

A hybrid CNN–Transformer architecture combining ResNet50 and Swin Transformer was developed to exploit complementary visual representations. ResNet50 captured local characteristics such as particle boundaries, texture, color variation, and fine morphological structures, whereas Swin Transformer modeled broader spatial relationships. Similar feature-fusion strategies have shown potential in recent microplastic analysis studies (Yao et al., 2025; Saur et al., 2026).

Feature vectors from the CNN and Transformer branches were projected into a common 256-dimensional latent space. Let fC and fT denote the projected CNN and Transformer feature vectors. A learnable gating module generated branch-specific scores sC and sT, which were normalized using softmax:

$$\alpha C = \exp(sC) / [\exp(sC) + \exp(sT)] \quad (5)$$

$$\alpha T = \exp(sT) / [\exp(sC) + \exp(sT)] \quad (6)$$

such that:

$$\alpha C + \alpha T = 1 \quad (7)$$

The fused representation was calculated as:

$$fFusion = \alpha C \times fC + \alpha T \times fT \quad (8)$$

This adaptive mechanism allows the relative contribution of local CNN features and global Transformer representations to vary among images. The fused representation was subsequently processed by layer normalization, dropout, a fully connected layer with 128 hidden units and GELU activation, followed by the final four-class classification layer.

Model training and performance evaluation

All deep-learning architectures were trained under an identical experimental protocol to ensure a fair comparison. A two-stage transfer-learning strategy was adopted. During the first stage, the pretrained backbone parameters were frozen and only the newly added classification, projection, and fusion layers were optimized for a maximum of five epochs. During the second stage, the upper layers of the pretrained backbones were unfrozen and fine-tuned for a maximum of 25 additional epochs using a lower learning rate. Transfer learning has been widely applied in microplastic image analysis, particularly when the available number of microscopy images is limited (Almeida et al., 2025; Liu et al., 2025; Harde and B, 2026).

Training was performed using the AdamW optimizer with a batch size of 16. The initial learning rate was 1×10^{-3} during the warm-up stage and 1×10^{-5} during fine-tuning, with a weight-decay coefficient of 1×10^{-4}. Weighted cross-entropy loss was employed to account for class imbalance, together with label smoothing of 0.03. A ReduceLROnPlateau scheduler reduced the learning rate when validation macro-F1 failed to improve, and early stopping was applied with a patience of five epochs. The model checkpoint achieving the highest validation macro-F1 was retained for evaluation.

For the conventional morphology–texture classifiers, five-fold stratified cross-validation was performed on the 909-image development subset, followed by final evaluation on the independent 228-image held-out test set. Deep-learning models were trained and validated using their fixed train/validation/test protocol, and final deep-learning performance was evaluated on the designated 171-image test subset. Model selection was primarily based on macro-F1.

Classification performance was quantified using accuracy, balanced accuracy, macro precision, macro recall, macro-F1, weighted-F1, Matthews correlation coefficient (MCC), Cohen's kappa, and one-versus-rest multiclass ROC-AUC. Class-specific precision, recall, and F1-score were additionally calculated for algae, filament, fragment, and pellet. Precision, recall, and F1-score are commonly used in automated microplastic classification and segmentation because they provide complementary information regarding false-positive and false-negative predictions (Phan et al., 2023).

$$\text{Accuracy} = \text{Number of correctly classified images} / \text{Total number of images} \quad (9)$$

$$\text{Precision}_k = TP_k / (TP_k + FP_k) \quad (10)$$

$$\text{Recall}_k = TP_k / (TP_k + FN_k) \quad (11)$$

$$F1_k = 2 \times (\text{Precision}_k \times \text{Recall}_k) / (\text{Precision}_k + \text{Recall}_k) \quad (12)$$

$$\text{Balanced Accuracy} = (1 / K) \times \Sigma\, \text{Recall}_k \quad (13)$$

$$\text{Macro-F1} = (1 / K) \times \Sigma\, F1_k \quad (14)$$

where TP_k, FP_k, and FN_k denote true-positive, false-positive, and false-negative predictions for class *k*, respectively, and K = 4 is the number of classes. The standard precision, recall, and F1 formulations are also explicitly reported in previous microplastic image-analysis work.

Raw and normalized confusion matrices were generated to identify systematic confusion among the four categories. One-versus-rest ROC curves and precision–recall curves were also calculated for class-wise analysis. Prediction-confidence distributions were retained to determine whether incorrect classifications were associated with systematically lower model confidence. The overall analytical workflow and proposed attention-gated CNN–Transformer architecture are summarized in Fig. 1.

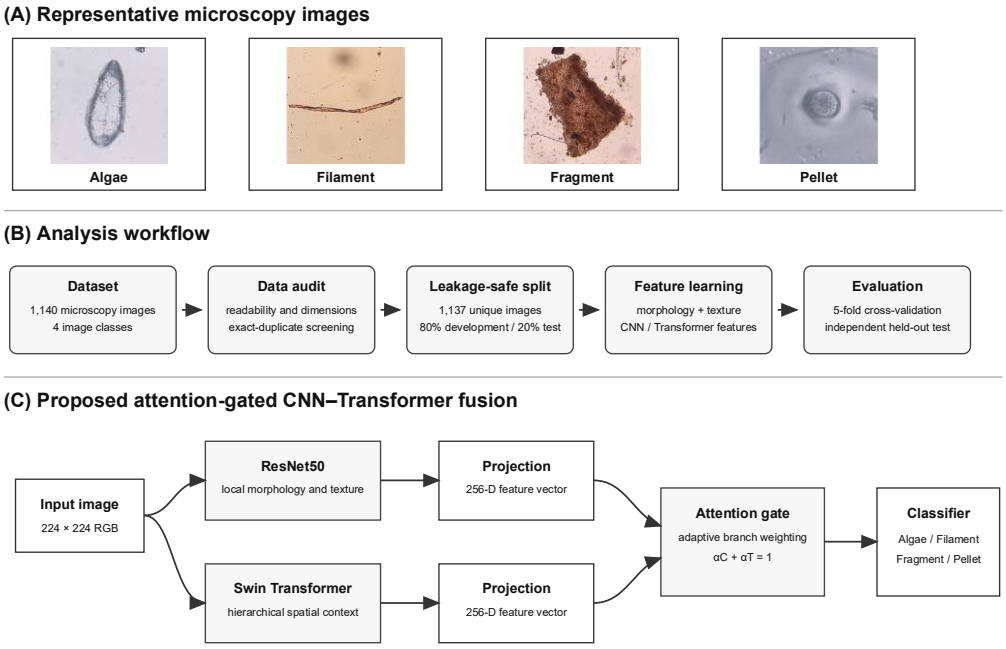


Fig. 1 Overall workflow of the proposed morphology-aware microplastic classification framework. (A) Representative optical microscopy images of algae, filament, fragment, and pellet classes. (B) Dataset auditing, preprocessing, morphology/texture analysis, deep-feature extraction, and model evaluation workflow. (C) Proposed attention-gated CNN–Transformer architecture combining ResNet50 and Swin Transformer representations for four-class classification

Statistical analysis and explainability

For the morphology–texture machine-learning models, stability was assessed through stratified five-fold cross-validation on the 909-image development subset, with performance summarized as mean ± standard deviation across folds. Deep-learning models were evaluated using the separate fixed train/validation/test protocol described above, and final performance metrics were calculated on the corresponding test subsets.

Interpretability of the conventional machine-learning analysis was examined using feature-importance rankings derived from the morphology and texture descriptors. For the CNN–Transformer fusion model, the contribution of

convolutional and Transformer representations was examined through the fusion mechanism, while particle-level interpretation was supported by morphology–texture feature analysis.

Results

Morphological characteristics of the image classes

After exact-duplicate removal, morphology analysis was performed on 1,137 unique microscopy images comprising 300 algae, 300 filament, 299 fragment, and 238 pellet images. Clear class-dependent differences were observed in the principal morphological descriptors.

Filaments exhibited the most elongated morphology, with a median aspect ratio of 5.657 and eccentricity of 0.984, compared with 2.122/0.882 for algae, 1.876/0.846 for fragments, and 1.154/0.499 for pellets. Pellets showed the highest median circularity (0.810) and solidity (0.986), whereas filaments had the lowest circularity (0.101) and solidity (0.505) and the highest boundary roughness (3.144). Median particle area fraction was 0.052 for filaments, 0.154 for algae, 0.195 for fragments, and 0.166 for pellets.

Kruskal–Wallis tests confirmed significant class effects for area fraction ($H = 398.229$, $p = 5.35 \times 10^{-86}$), aspect ratio and eccentricity ($H = 705.349$, $p = 1.45 \times 10^{-152}$), circularity and boundary roughness ($H = 511.340$, $p = 1.66 \times 10^{-110}$), and solidity ($H = 388.354$, $p = 7.38 \times 10^{-84}$). These results demonstrate substantial differences among classes in elongation, compactness, occupied area, and boundary complexity. The distributions of these morphological descriptors are shown in Fig. 2.

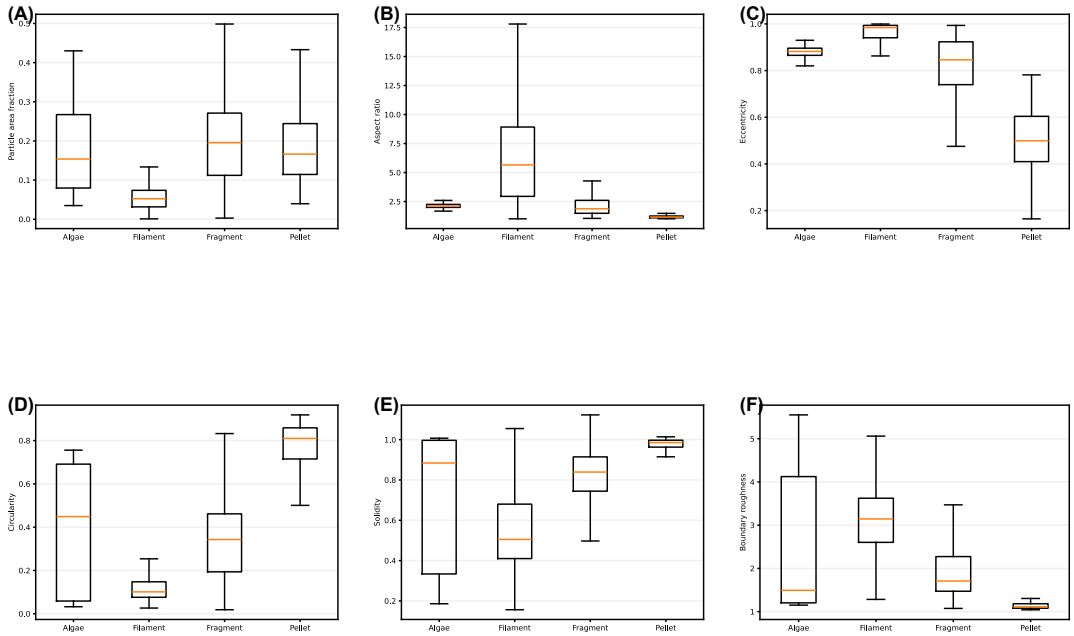


Fig. 2 Distribution of major particle-level morphological descriptors across algae, filament, fragment, and pellet classes: (A) particle area fraction, (B) aspect ratio, (C) eccentricity, (D) circularity, (E) solidity, and (F) boundary roughness index. All analyses were performed on the duplicate-free dataset of 1,137 unique microscopy images

Texture, feature-space, and feature-importance analysis

Texture and structural-complexity descriptors differed significantly among the four classes. GLCM contrast showed a significant class effect ($H = 120.141$, $p = 7.20 \times 10^{-26}$), while fractal dimension exhibited an even stronger difference ($H = 443.268$, $p = 9.37 \times 10^{-96}$).

Principal component analysis of the standardized 17-feature morphology–texture representation showed that PC1 and PC2 explained 41.13% and 14.76% of the total variance, respectively. The feature space showed partial class separation, particularly for elongated filaments and compact pellets, whereas algae and fragments exhibited greater overlap.

Random Forest feature-importance analysis identified aspect ratio (0.1700) and eccentricity (0.1679) as the strongest predictors, followed by extent (0.0885), circularity (0.0689), boundary roughness (0.0638), solidity (0.0567), grayscale mean (0.0549), convexity (0.0538), mean skeleton width (0.0480), and edge density (0.0392). Overall, morphology-related descriptors contributed most strongly to class discrimination, with texture and intensity features providing complementary information. The corresponding texture, feature-space, and feature-importance analyses are shown in Fig. 3.

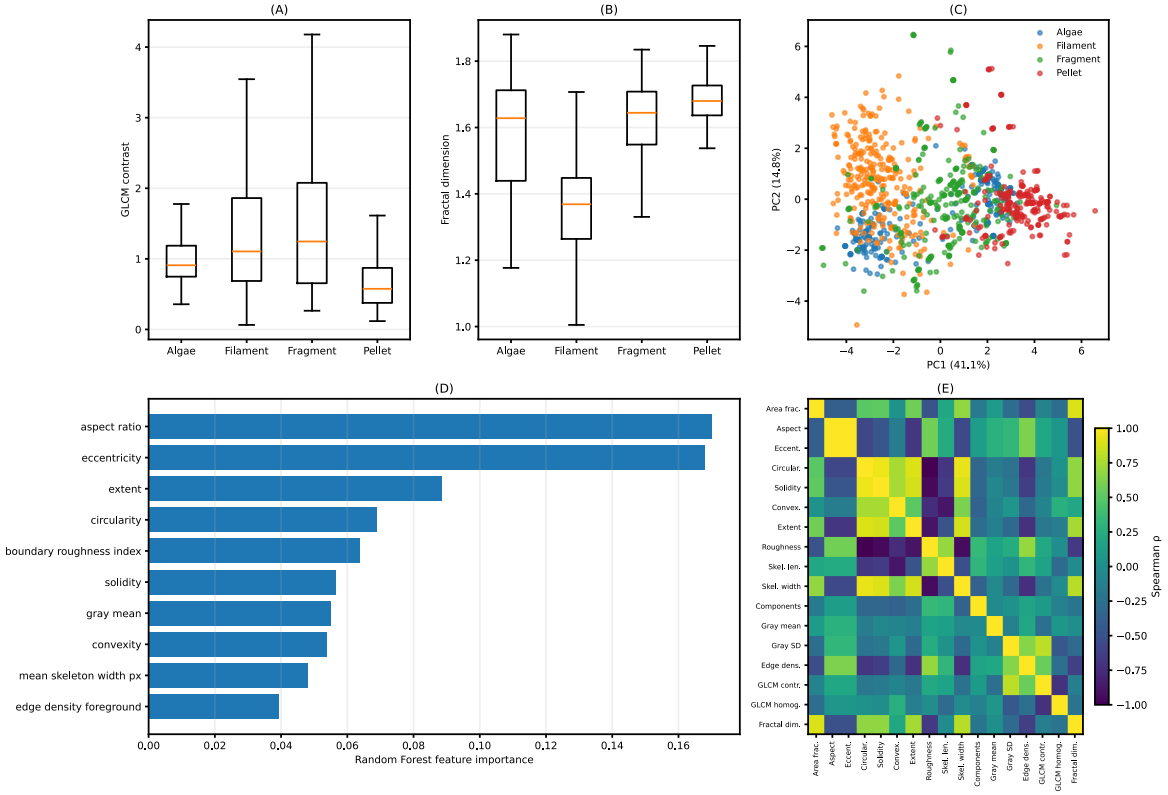


Fig. 3 Texture, multivariate feature-space, and feature-importance analysis of the duplicate-free microscopy dataset. (A) Class-wise distribution of GLCM contrast; (B) fractal dimension; (C) principal component analysis of the standardized 17-dimensional morphology–texture feature set; (D) ten highest Random Forest feature-importance scores; and (E) Spearman correlation matrix of the extracted descriptors. PC1 and PC2 explained 41.13% and 14.76% of the total variance, respectively

Machine-learning classification performance

The four conventional machine-learning models were compared using five-fold stratified cross-validation on the 909-image development subset, while the remaining 228 images were retained as an independent held-out test set.

Among the evaluated classifiers, HistGradientBoosting achieved the highest cross-validation performance, with a mean accuracy of 0.9329 ± 0.0253, balanced accuracy of 0.9344 ± 0.0250, and macro-F1 of 0.9341 ± 0.0255. RBF-

SVM ranked second with a macro-F1 of 0.9298 ± 0.0189, followed by Random Forest (0.9248 ± 0.0184) and Extra Trees (0.9205 ± 0.0237).

The cross-validation results for the four machine-learning classifiers are summarized in Table 2.

Table 2 Five-fold cross-validation performance of the machine-learning classifiers on the development set

Model	Accuracy	Balanced Accuracy	Macro-F1	MCC
HistGradientBoosting	**0.9329 ± 0.0253**	**0.9344 ± 0.0250**	**0.9341 ± 0.0255**	**0.9105 ± 0.0338**
RBF-SVM	0.9296 ± 0.0190	0.9317 ± 0.0189	0.9298 ± 0.0189	0.9064 ± 0.0254
Random Forest	0.9230 ± 0.0174	0.9250 ± 0.0175	0.9248 ± 0.0184	0.8975 ± 0.0233
Extra Trees	0.9186 ± 0.0224	0.9209 ± 0.0217	0.9205 ± 0.0237	0.8919 ± 0.0295

Note. Values are reported as mean ± standard deviation across five stratified folds of the 909-image development subset. The highest value for each metric is shown in bold.

Accordingly, HistGradientBoosting was selected as the best morphology–texture classifier and evaluated on the untouched test set. The model achieved an accuracy of 0.9342, balanced accuracy of 0.9333, macro-F1 of 0.9353, MCC of 0.9124, Cohen's κ of 0.9120, and a macro one-versus-rest ROC-AUC of 0.9944. Detailed held-out test performance is reported in Table 3.

Class-wise performance remained consistently high. Filaments achieved the highest F1-score (0.9580), followed by pellets (0.9462), algae (0.9412), and fragments (0.8960). Fragment classification was comparatively more challenging, primarily because of reduced precision (0.8615) despite a recall of 0.9333. The cross-validation and held-out test performance are illustrated in Fig. 4.

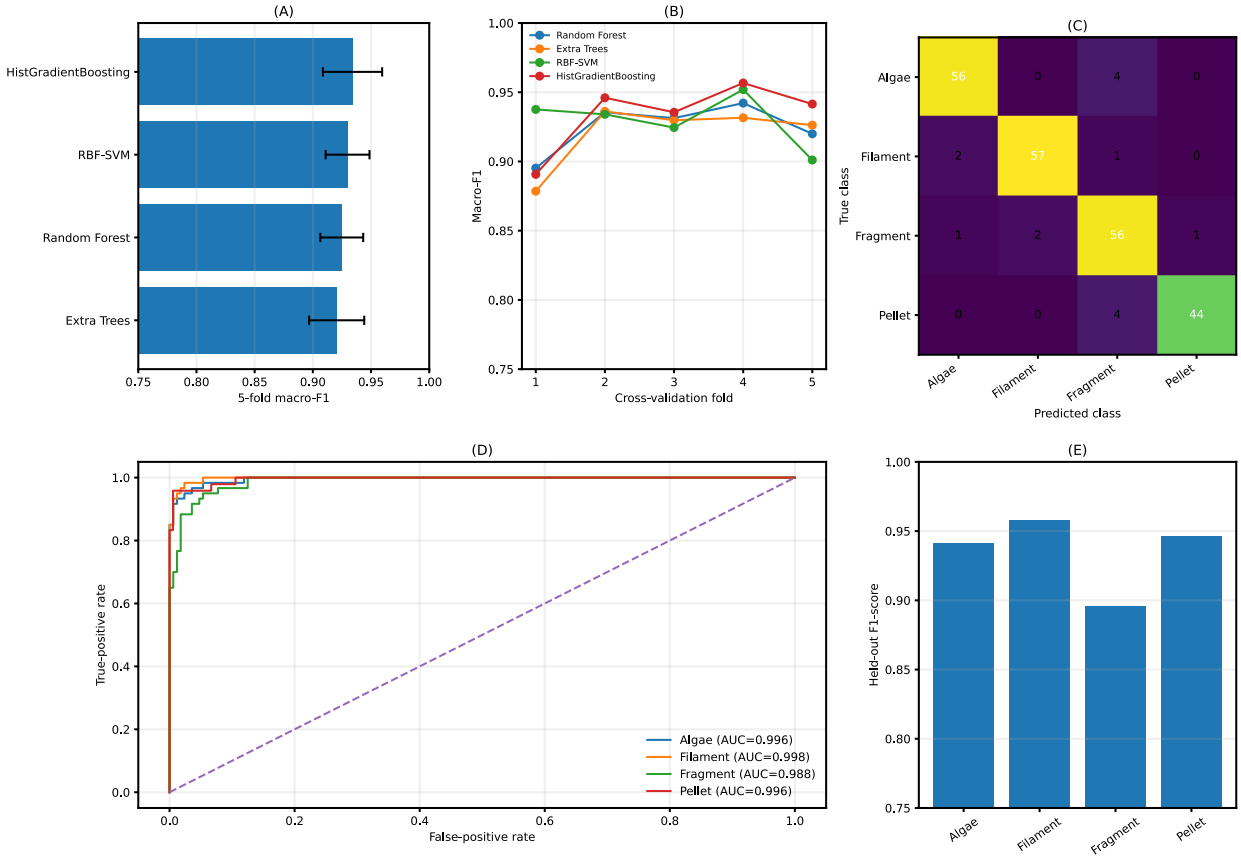


Fig. 4 Performance of morphology–texture machine-learning classifiers under the leakage-controlled evaluation protocol. (A) Mean five-fold macro-F1 with standard deviation for Random Forest, Extra Trees, RBF-SVM, and HistGradientBoosting; (B) fold-wise macro-F1 stability; (C) confusion matrix of the selected HistGradientBoosting model on the independent held-out test set; (D) class-specific one-versus-rest ROC curves; and (E) held-out class-wise F1-scores. Model comparison was performed using the 909-image development subset, and final performance was evaluated on the untouched 228-image test subset

Table 3 Held-out test performance of the selected HistGradientBoosting classifier

Class / Metric	Precision	Recall	F1-score	Support

Algae	0.9492	0.9333	0.9412	60
Filament	0.9661	0.9500	**0.9580**	60
Fragment	0.8615	0.9333	0.8960	60
Pellet	0.9778	0.9167	0.9462	48
Overall accuracy	—	—	**0.9342**	**228**
Balanced accuracy	—	—	**0.9333**	—
Macro-F1	—	—	**0.9353**	—
MCC	—	—	**0.9124**	—
Cohen's κ	—	—	**0.9120**	—
Macro ROC-AUC (OvR)	—	—	**0.9944**	—

Note. Results were obtained exclusively from the independent 228-image held-out test set, which was not used for model fitting or model selection.

Prediction confidence and error analysis

Prediction-confidence analysis revealed differences between correct and incorrect classifications. The mean maximum predicted probability was 0.9940 for correctly classified images and 0.9067 for misclassified images, while the median top-1 versus top-2 probability margin was 1.0000 and 0.9531, respectively.

Among the 228 held-out images, 15 were misclassified (6.58%), and 12 of these errors had prediction confidence ≥ 0.90. Using five equal-frequency confidence bins, the top-label expected calibration error was 0.0540, indicating a moderate discrepancy between confidence and empirical correctness.

The most frequent errors were algae predicted as fragments (n = 4) and pellets predicted as fragments (n = 4). Other error pathways were less frequent and are summarized in Table 4, indicating that fragments represented the principal source of classification ambiguity. Prediction-confidence distributions and error patterns are illustrated in Fig. 5.

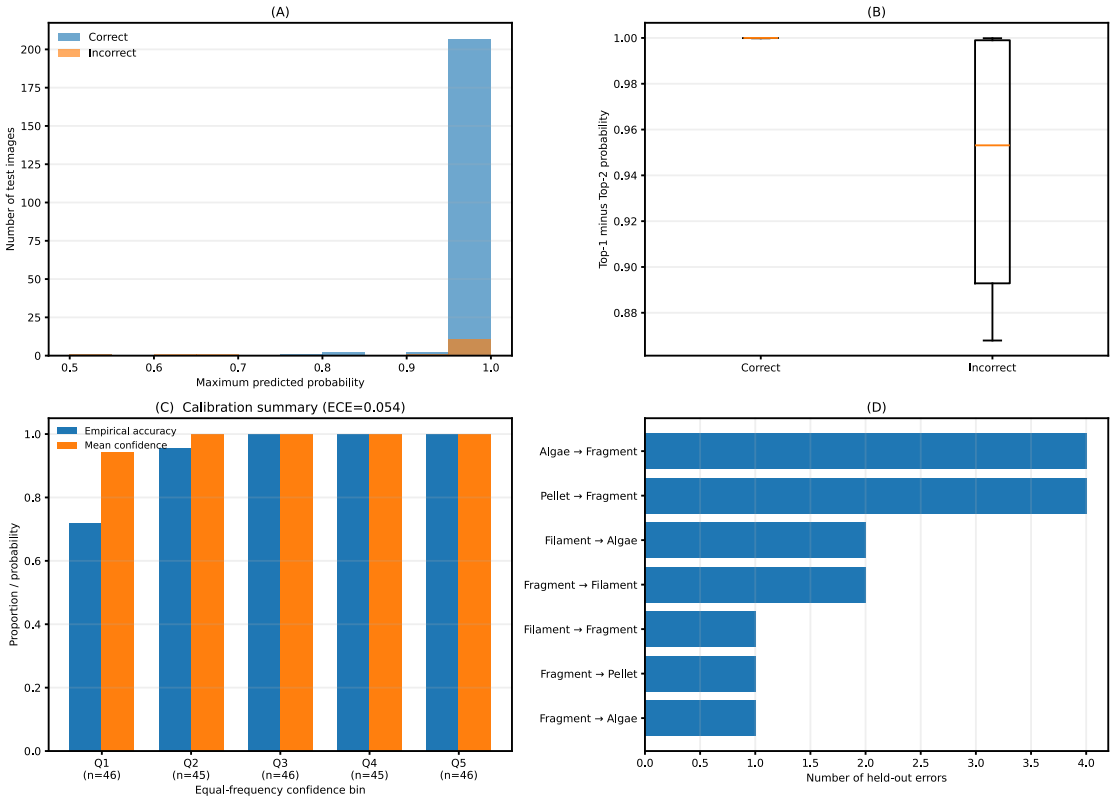


Fig. 5 Prediction-confidence and error analysis of the selected HistGradientBoosting classifier on the independent held-out test set. (A) Distribution of maximum predicted probabilities for correctly and incorrectly classified images; (B) top-1 versus top-2 probability margins; (C) calibration summary using five equal-frequency confidence bins, with an expected calibration error (ECE) of 0.0540; and (D) observed class-to-class misclassification patterns. Fifteen of 228 test images were incorrectly classified

Table 4 Misclassification patterns on the independent held-out test set

True class	Predicted class	Number of errors	Mean prediction confidence
Algae	Fragment	**4**	0.8949
Pellet	Fragment	**4**	0.8773
Fragment	Filament	2	0.9872
Filament	Algae	2	0.9783
Fragment	Pellet	1	0.9997
Filament	Fragment	1	0.9288
Fragment	Algae	1	0.6517
Total	—	**15**	—

Note. Mean prediction confidence represents the average maximum class probability among images belonging to each specific misclassification pathway.

Deep-learning and CNN–Transformer fusion performance

Three deep-learning architectures were evaluated on the designated 171-image test subset: ResNet50 as the CNN baseline, Swin Transformer as the hierarchical attention-based model, and the proposed attention-gated CNN–Transformer fusion framework combining ResNet50 and Swin Transformer.

ResNet50 achieved an accuracy of 0.9064, macro-F1 of 0.9035, and macro-AUC of 0.9880. Swin Transformer improved performance to an accuracy of 0.9240, macro-F1 of 0.9216, and macro-AUC of 0.9912. The proposed Attention Fusion model achieved an accuracy of 0.9240, macro-F1 of 0.9222, and macro-AUC of 0.9954, while precision and recall were 0.9225 and 0.9220, respectively.

As summarized in Table 5, Attention Fusion and Swin Transformer achieved identical accuracy (0.9240), whereas the fusion model provided slightly higher macro-F1 and macro-AUC values, indicating the benefit of combining complementary CNN and Transformer representations.

Table 5 Performance comparison of the evaluated deep-learning models

Model	Accuracy	Precision	Recall	Macro-F1	Macro-AUC
ResNet50	0.9064	0.9103	0.8991	0.9035	0.9880
Swin Transformer	0.9240	0.9222	0.9215	0.9216	0.9912
Attention Fusion	**0.9240**	**0.9225**	**0.9220**	**0.9222**	**0.9954**

Error analysis and model interpretation of the Attention Fusion framework

Class-wise analysis showed balanced performance across the four categories. Algae achieved the highest F1-score (0.97), followed by pellet (0.94), filament (0.91), and fragment (0.87). The comparatively lower performance for fragments is consistent with their irregular morphology and visual similarity to other particle categories (Table 6).

The confusion matrix showed that most samples were correctly classified, with the remaining errors occurring mainly among morphologically similar categories. The confusion matrix, multi-class ROC curves, and class-wise F1-scores of the Attention Fusion model are illustrated in Figure 6(D–F).

Class-wise ROC-AUC values were 1.0000 for algae, 0.9950 for filament, 0.9886 for fragment, and 0.9982 for pellet, yielding a macro-AUC of 0.9954 (Table 7). These results demonstrate strong discrimination across all four classes.

Table 6 Class-wise classification performance of the Attention Fusion model

Class	Precision	Recall	F1-score	Support
algae	0.96	0.98	0.97	51
filament	0.91	0.91	0.91	46
fragment	0.88	0.86	0.87	42
pellet	0.94	0.94	0.94	32
Accuracy	—	—	**0.92**	**171**
Macro Average	0.92	0.92	0.92	171

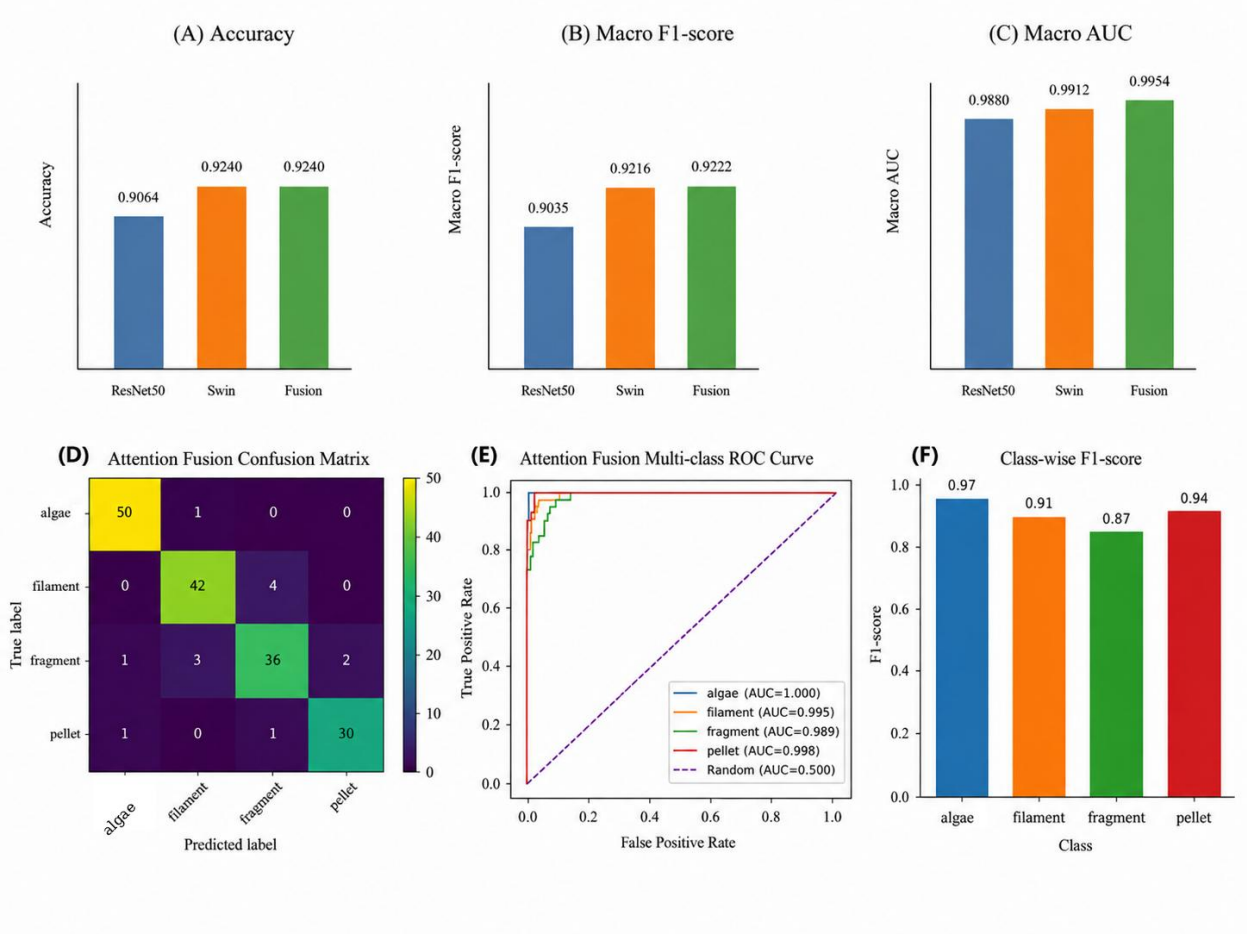


Fig. 6 Deep-learning model comparison and performance analysis of the Attention Fusion framework. (A) Accuracy comparison; (B) Macro F1-score comparison; (C) Macro AUC comparison of ResNet50, Swin Transformer, and Attention Fusion models; (D) confusion matrix of the Attention Fusion model; (E) multi-class ROC curves; and (F) class-wise F1-score comparison for algae, filament, fragment, and pellet classes

Table 7 Class-wise ROC-AUC performance of the Attention Fusion model

Class	**AUC**
algae	1.0000
filament	0.9950
fragment	0.9886
pellet	0.9982
Macro AUC	**0.9954**

Discussion

Among the evaluated deep-learning architectures, ResNet50 achieved an accuracy of 0.9064 and macro-F1 of 0.9035, while Swin Transformer improved accuracy to 0.9240 and macro-AUC to 0.9912. The proposed Attention Fusion framework achieved the most balanced deep-learning performance, with an accuracy of 0.9240, macro-F1 of 0.9222, and macro-AUC of 0.9954. These findings support the complementary use of CNN-based local feature extraction and Transformer-based contextual representation for heterogeneous microscopy images.

Class-wise analysis showed stronger recognition of algae and pellet samples, whereas fragments were comparatively more difficult to classify. This likely reflects the irregular morphology and visual overlap of fragments with other particle categories. The use of field-collected rather than artificially generated particles increases the environmental relevance of the evaluation and highlights the challenges associated with naturally heterogeneous samples.

Conclusions

This study presented an automated deep-learning framework for classification of microscopy-based microplastic images using convolutional and Transformer-based feature representations. A comprehensive evaluation was performed using three architectures, including ResNet50, Swin Transformer, and the proposed Attention Fusion model.

The experimental results demonstrated that deep-learning approaches effectively recognize different microplastic categories. Among the evaluated deep-learning architectures, the proposed Attention Fusion framework achieved the most balanced performance, obtaining an accuracy of 0.9240, a macro-F1 score of 0.9222, and a macro-AUC value of 0.9954. The results indicate that integrating local morphological information from CNN features with global contextual information from Transformer attention mechanisms can improve class-level discrimination and classification robustness.

Although the model achieved strong performance on the evaluated dataset, further validation using independent datasets from different geographic regions, imaging conditions, and environmental matrices is required to assess its generalization capability.

Author contributions

Majid Hajibarati: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Visualization, Writing – original draft, Writing – review & editing.

Statements and Declarations

Ethical responsibilities of authors

All authors have read, understood, and have complied as applicable with the statement on “Ethical responsibilities of Authors” as found in the Instructions for Authors.

Ethics approval

Not applicable. This study used a publicly available microscopy image dataset of environmental microplastic particles and diatom algae and did not involve human participants, human data, human tissue, or animals.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The author declares no competing interests.

Funding

No funding was received for conducting this study or preparing this manuscript.

Data availability

The microscopy image dataset analyzed in this study is publicly available in Mendeley Data under the title “Microplastic images and Automatic classification by deep learning,” Version 1, https://doi.org/10.17632/yyn2g83vsy.1. No new primary image dataset was generated in the present study.

References

Abimbola, I., Thevar, T., McAfee, M., Creedon, L., Khosravi, H., & Gharbia, S. (2025). Holographic imaging and machine learning for microplastic size and shape analysis in water. *Emerging Contaminants*, *11*, 100558. https://doi.org/10.1016/j.emcon.2025.100558

Almeida, V., Wiecheteck, G.K., Christo, S.W., Girard, P., de Souza, J.B., Inglez, J.E.F., Staichak, G., & Ferreira Júnior, A.L. (2025). Application of a convolutional neural network for automated multiclass identification of field-collected microplastics and diatom algae from optical microscopy images. *Revista Brasileira de Ciências Ambientais*, *60*, e2491. https://doi.org/10.5327/Z2176-94782491

Bianco, V., Memmolo, P., Carcagnì, P., Merola, F., Paturzo, M., Distante, C., & Ferraro, P. (2020). Microplastic identification via holographic imaging and machine learning. *Advanced Intelligent Systems*, *2*(2), 1900153. https://doi.org/10.1002/aisy.201900153

Cowger, W., Gray, A., Christiansen, S.H., DeFrond, H., Deshpande, A.D., Hemabessiere, L., Lee, E., Mill, L., Munno, K., Ossmann, B.E., Pittroff, M., Rochman, C., Sarau, G., Tarby, S., & Primpke, S. (2020). Critical review of processing and classification techniques for images and spectra in microplastic research. *Applied Spectroscopy*, *74*(9), 989–1010. https://doi.org/10.1177/0003702820929064

Fernandez, G., Vito, D., Kaziz, K., Booth, D., Suresh-Babu, S., Shelke, S.S., & Banni, M. (2026). Advancing marine microplastic monitoring through deep learning-based image segmentation. *One Health Advances*, *4*, 24. https://doi.org/10.1186/s44280-026-00120-7

Giardino, M., Balestra, V., Janner, D., & Bellopede, R. (2023). Automated method for routine microplastic detection and quantification. *Science of the Total Environment*, *859*, 160036. https://doi.org/10.1016/j.scitotenv.2022.160036

Goyetche, R., Kortazar, L., & Amigo, J.M. (2023). Issues with the detection and classification of microplastics in marine sediments with chemical imaging and machine learning. *TrAC Trends in Analytical Chemistry*, *166*, 117221. https://doi.org/10.1016/j.trac.2023.117221

Harde, N., & B, T. (2026). Deep learning based classification of microplastic in edible food using optical microscopy images. *Discover Artificial Intelligence*, *6*, 478. https://doi.org/10.1007/s44163-026-01246-8

Höppener, E.M., Shahmohammadi, S., Parker, L.A., Henke, S., & Urbanus, J.H. (2023). Classification of (micro)plastics using cathodoluminescence and machine learning. *Talanta*, *253*, 123985. https://doi.org/10.1016/j.talanta.2022.123985

Huang, H., Cai, H., Qureshi, J.U., Mehdi, S.R., Song, H., Liu, C., Di, Y., Shi, H., Yao, W., & Sun, Z. (2023). Proceeding the categorization of microplastics through deep learning-based image segmentation. *Science of the Total Environment*, *896*, 165308. https://doi.org/10.1016/j.scitotenv.2023.165308

Kim, J.Y., Koh, E.H., Yang, J.Y., Mun, C., Lee, S., Lee, H., Kim, J., Park, S.G., Kang, M., Kim, D.H., & Jung, H.S. (2024). 3D plasmonic gold nanopocket structure for SERS machine learning-based microplastic detection. *Advanced Functional Materials*, *34*(2), 2307584. https://doi.org/10.1002/adfm.202307584

Lee, G., & Jhang, H. (2021). Neural network analysis for microplastic segmentation. *Sensors*, *21*(21), 7030. https://doi.org/10.3390/s21217030

Lee, G., Jung, J., Moon, S., Jung, J., & Jhang, K. (2024). Microscopic image dataset with segmentation and detection labels for microplastic analysis in sewage: Enhancing research and environmental monitoring. *Microplastics*, *3*, 264–275. https://doi.org/10.3390/microplastics3020016

Li, Y., Zhu, Y., Huang, J., Ho, Y.W., Fang, J.K.H., & Lam, E.Y. (2024). High-throughput microplastic assessment using polarization holographic imaging. *Scientific Reports*, *14*, 2355. https://doi.org/10.1038/s41598-024-52762-5

Liu, Y., Zhao, G., & Liu, F. (2025). Deep learning-based shape classification for hyperspectral-imaged microplastics. *Analytical Chemistry*, *97*(38), 20798–20808. https://doi.org/10.1021/acs.analchem.5c02683

Lorenzo-Navarro, J., Castrillón-Santana, M., Sánchez-Nielsen, E., Zarco, B., Herrera, A., Martínez, I., & Gómez, M. (2021). Deep learning approach for automatic microplastics counting and classification. *Science of the Total Environment*, *765*, 142728. https://doi.org/10.1016/j.scitotenv.2020.142728

Maes, T., Jessop, R., Wellner, N., Haupt, K., & Mayes, A.G. (2017). A rapid-screening approach to detect and quantify microplastics based on fluorescent tagging with Nile Red. *Scientific Reports*, *7*, 44501. https://doi.org/10.1038/srep44501

Massarelli, C., Campanale, C., & Uricchio, V.F. (2021). A handy open-source application based on computer vision and machine learning algorithms to count and classify microplastics. *Water*, *13*(15), 2104. https://doi.org/10.3390/w13152104

Meyers, N., Catarino, A.I., Declercq, A., Brenan, A., Devriese, L., Vandegehuchte, M., De Witte, B., Janssen, C., & Everaert, G. (2022). Microplastic detection and identification by Nile red staining: Towards a semi-automated, cost- and time-effective technique. *Science of the Total Environment*, *823*, 153441. https://doi.org/10.1016/j.scitotenv.2022.153441

Morét-Ferguson, S., Law, K.L., Proskurowski, G., Murphy, E.K., Peacock, E.E., & Reddy, C.M. (2010). The size, mass, and composition of plastic debris in the western North Atlantic Ocean. *Marine Pollution Bulletin*, *60*(10), 1873–1878. https://doi.org/10.1016/j.marpolbul.2010.07.020

Park, H.M., Park, S., de Guzman, M.K., Baek, J.Y., Cirkovic Velickovic, T., Van Messem, A., & De Neve, W. (2022). MP-Net: Deep learning-based segmentation for fluorescence microscopy images of microplastics isolated from clams. *PLOS ONE*, *17*(6), e0269449. https://doi.org/10.1371/journal.pone.0269449

Phan, S., Torrejon, D., Furseth, J., Mee, E., & Luscombe, C. (2023). Exploiting weak supervision to facilitate segmentation, classification, and analysis of microplastics (<100 μm) using Raman microspectroscopy images. *Science of the Total Environment*, *886*, 163786. https://doi.org/10.1016/j.scitotenv.2023.163786

Prata, J.C., Alves, J.R., da Costa, J.P., Duarte, A.C., & Rocha-Santos, T. (2020). Major factors influencing the quantification of Nile Red stained microplastics and improved automatic quantification (MP-VAT 2.0). *Science of the Total Environment*, *719*, 137498. https://doi.org/10.1016/j.scitotenv.2020.137498

Rermborirak, K., Nanuan, P., Komonpan, P., & Sukpancharoen, S. (2025). Low-cost portable microplastic detection system integrating Nile Red fluorescence staining with YOLOv8-based deep learning. *Journal of Hazardous Materials Advances*, *19*, 100787. https://doi.org/10.1016/j.hazadv.2025.100787

Royer, S.J., Wolter, H., Delorme, A.E., Lebreton, L., & Poirion, O.B. (2024). Computer vision segmentation model—deep learning for categorizing microplastic debris. *Frontiers in Environmental Science*, *12*, 1386292. https://doi.org/10.3389/fenvs.2024.1386292

Salehin, S., Ahmed, S.S.U., Tefek, U., Weirauch, L., & Hanay, M.S. (2026). Microwave cytometry with machine learning for shape-resolved microplastic detection. *ACS Sensors*, *11*(4), 3105–3119. https://doi.org/10.1021/acssensors.5c04073

Sarker, M.A.B., Imtiaz, M.H., Holsen, T.M., & Baki, A.B.M. (2024). Real-time detection of microplastics using an AI camera. *Sensors*, *24*(13), 4394. https://doi.org/10.3390/s24134394

Saur, L., von Pawlowski, M., Gengenbach, U., Sieber, I., Shirali, H., Wührl, L., Weng, X., Kiko, R., & Pylatiuk, C. (2026). Classification of microplastic particles in water using polarized light scattering and machine learning methods. *Journal of Hazardous Materials Advances*, *23*, 101343. https://doi.org/10.1016/j.hazadv.2026.101343

Scheiterlein, T., & Fiener, P. (2026). Microplastics detection in agricultural soil combining 3D Laser Scanning Confocal Microscopy with machine learning. *Microplastics and Nanoplastics*, *6*, 47. https://doi.org/10.1186/s43591-026-00202-8

Shi, B., Patel, M., Yu, D., Yan, J., Li, Z., Petriw, D., Pruyn, T., Smyth, K., Passeport, E., Miller, R.J.D., & Howe, J.Y. (2022). Automatic quantification and classification of microplastics in scanning electron micrographs via deep learning. *Science of the Total Environment*, *825*, 153903. https://doi.org/10.1016/j.scitotenv.2022.153903

Smolen, J.A., Moore, G.E., Perez, N.D., & Wooley, K.L. (2025). Adaptable microplastic classification using similarity learning on μFTIR spectra collected from μFTIR focal plane array imaging. *Proceedings of the National Academy of Sciences of the United States of America*, *122*(42), e2509745122. https://doi.org/10.1073/pnas.2509745122

Tsuchiya, M., Kitahashi, T., Taira, Y., Saito, H., Oguri, K., Nakajima, R., Lindsay, D.J., & Fujikura, K. (2025). Rapid detection and quantification of Nile Red-stained microplastic particles in sediment samples. *PeerJ*, *13*, e19196. https://doi.org/10.7717/peerj.19196

Vitali, C., Peters, R.J.B., Janssen, H.G., Undas, A.K., Munniks, S., Ruggeri, F.S., & Nielen, M.W.F. (2024). Quantitative image analysis of microplastics in bottled water using artificial intelligence. *Talanta*, *266*, 124965. https://doi.org/10.1016/j.talanta.2023.124965

Xu, J., & Wang, Z. (2024). Efficient and accurate microplastics identification and segmentation in urban waters using convolutional neural networks. *Science of the Total Environment*, *911*, 168696. https://doi.org/10.1016/j.scitotenv.2023.168696

Yang, C., Lin, X., Xu, J.L., Ferreira, R.O.G., Van de Wiele, T., Marinelli, L., Canga, E.M., & Gowen, A.A. (2026). Deep learning-assisted identification and quantification of cell-associated microplastics using darkfield hyperspectral imaging. *Talanta*, *308*, 129864. https://doi.org/10.1016/j.talanta.2026.129864

Yao, Y., Xu, W., & Fan, H. (2025). A deep learning approach for microplastic segmentation in microscopic images. *Toxics*, *13*(12), 1018. https://doi.org/10.3390/toxics13121018