Final Report

Group 9

Title: **Deep Learning for Microplastic Classification**

Team Members:

1. Anugya Khadka

2. Jhansi Balla

3. Nikhitha Rayabarapu

4. Priyanka Chowdary Mannava

Abstract:

The project will be focused on automation of microplastic particle recognition and classification in microscopy images by applying computer vision technologies to images. Its DeepParticle (MICRO) dataset, which was preprocessed with the help of cropping, normalization and comprehensive data augmentation, has six particle classes. Classical machine learning models which is Random Forest Algorithm and SVM along with deep learning models which are VGG16, ResNet-50, Vision Transformer were run to compare them with each other. SMOTE oversampling was used to counter the imbalance of classes. Observing the result from these experiments the fine-tuned ViTs got better accuracy of 98% as compared to other models. Grad-CAM visualizations and data augmentation and transfer learning were shown to enhance model robustness and give information about class-discriminative regions. The results prove that deep learning methods, especially transformers, provide a highly scalable, explainable, and powerful solution to environmental microplastic monitoring and are more accurate and generalize better than handcrafted feature-based methods.

Introduction:

Microplastics (less than 5 mm) have become one of the most ubiquitous pollutants of the marine and freshwater ecosystem. Their concentration endangers aquatic organisms, pollutes food chains and long-term environmental and health hazards. Conventionally, microscopic identification of microplastics is a tedious, laborious and inaccurate procedure. Their differences in particle size, color and texture render them difficult to identify in a consistent manner even when using trained analytical personnel. Therefore, automated and data-driven approaches are needed to aid in large-scale environmental monitoring and hasten plastic pollution research.

Research Question:

The proposed project will examine the hypothesis of whether deep learning models are more effective than classical machine learning methods in the correct detection and classification of microplastic particles in microscopy images. This work is directed by the following empirical question:

Do deep learning models, and specifically convolutional neural networks (CNNs) and Vision Transformers (ViTs), have a substantial effect on the accuracy and generalization of microplastic detection in comparison to the classifiers like SVM and Random Forest?

Literature Review:

The microplastics identification based on this paper, through convolutional neural network, integrated the micro-Raman spectroscopy data with a CNN used to identify 10 reference microplastic samples and three environmental samples (Ren et al., 2023). Here the CNN got the accuracy of approx. 96% on spectra and 95% on environmental samples.

The dual-modality spectral and image data were used to provide the accuracy of automatic microplastic classification and combined μFTIR spectral data and image data of 5 types of polymers, and models, such as AlexNet, ResNet18, ViT were compared (Sukkuea et al., 2025). AlexNet+Logistic Regression had a validation accuracy of approximately 99.03 and test of approximately 99.99.

The images in this paper were obtained using scanning electron microscopy (SEM) of microplastic particles and semantic segmentation models (U-Net, MultiResUNet) and then a finetuned CNN (VGG16) was used on the image to classify the shape with an accuracy of about 98.33% (Shi et al., 2022). It tackles the difficulty of dense and entangled fibres and small particles sizes by means of instance segmentation and object-focused embedding.

Proposed in this paper is a hybrid model of CNN (local spectral features) and Transformer (long-range dependencies) on FTIR spectral data of microplastics in soil, air, sediment and water, which is able to reach an accuracy of approximately 95.77 (Li et al., 2025). Although this applies spectral data, the architecture (CNN + Transformer) is similar to what you are interested in comparing CNNs and transformer-based networks (ViTs). It demonstrates that the components of transformers can be used in the classification of microplastic.

Methods and Experimental Design:

Dataset: DeepParticle Dataset (MICRO)

The DeepParticle (MICRO) dataset with six classes - foam, hard, line, noise, pellet, and reference was processed by cropping particles from annotation ".tsv" files and saving them in processed_MICRO. Images were resized, normalized, and augmented (flips, rotation, brightness/contrast, Gaussian noise) using Albumentations, then split 70/15/15 for train, validation, and test. The dataset is organized with the following approximate distribution:

Split	# Images
Train	3547
Validation	760
Test	761

Data Characteristics:

- Images are in RGB Format.
- Classes are imbalanced, particularly the "hard" class which is the majority.

Data preprocessing:

- Image Transformations:

  <u>Training Set:</u> Includes RandomResizedCrop (224×224), horizontal/vertical flip, rotation (±25°), brightness/contrast changes, Gaussian noise, and normalization.

Validation/test set: Resize to 224×224 with normalization only.

- SMOTE Oversampling:
  Due to the extreme class imbalance, SMOTE (Synthetic Minority Over-sampling Technique) was used. SMOTE extracted features using flattened color histograms (8×8×8 bins per RGB channel). A balanced training set was created by randomly re-mapping oversampled images to true image pathways.

Applying SMOTE to balance training data...

Preparing SMOTE features: 100%|██████████████████████████████|
3547/3547 [00:00<00:00, 5161.51it/s]

Original samples: 3547, After SMOTE: 14874

Reproducibility Practices:

- Random seeds (np.random.seed(42) and torch.manual_seed(42)) are fixed.
- CUDA deterministic behavior is used where possible.

Code and library versions (standard versions in colab):

PyTorch==2.1, Albumentations==1.2, scikit-learn==1.3, imbalanced-learn==0.11.

Experimental Setup:

Batch size: 32 (train/val), with 1 test for visualization.

Epochs: up to 100.

Early Stopping: 10 epochs

Metrics: Accuracy, precision, recall, F1 score, per-class IOU Scores, Confusion matrix.

Explainability Factor: Grad-CAM was applied to the final convolutional layer (model.features[-1]) to visualize class-discriminative regions.

Baselines: ML: Random Forest, SVM; CNN: VGG16, Resnet-50; Transformers: ViTs

Model Architecture:

Model	**Type**	**Implementation Summary**
Random Forest Classifier	Classical ML	200-tree ensemble using handcrafted features (GLCM texture, intensity statistics, entropy, edges).
Support Vector Machine (SVM)	Classical ML	RBF kernel (C=10, gamma='scale'), trained on same handcrafted features.
VGG16 (Fine-tuned)	Deep Learning (CNN)	Pretrained VGG16 from ImageNet. Feature extractor layers frozen; fully connected head replaced with Linear(4096 → 6) for six classes. Trained with Adam (LR=1e-4),

		CrossEntropyLoss, early stopping, and learning-rate scheduling.
Resnet-50	Deep Learning (CNN)	Pretrained on ImageNet. Backbone frozen. Fully connected layer replaced with: Linear(in_features → 512) → ReLU → Dropout(0.4) → Linear(512 → num_classes) Optimizer: Adam, LR=1e-4, CrossEntropyLoss. ReduceLROnPlateau scheduler (factor=0.5, patience=3).
ViTs	Deep Learning (Transformers)	Pretrained ViT-B/16, encoder frozen, classification head replaced with Linear(embed_dim → num_classes), Adam (LR=1e-4), CrossEntropyLoss, ReduceLROnPlateau, SMOTE-balanced training set, early stopping applied.

<u>Results And Analysis:</u>

Performance Comparision:

Model	Epochs	Accuracy	Precision	Recall	F1 Score
RF	-	0.87	0.84	0.710	0.748
SVM	-	0.87	0.85	0.718	0.741
Resnet-50	31	0.93	0.88	0.96	0.92
VGG	19	0.97	0.96	0.96	0.96
ViTs	30	**0.98**	**0.97**	**0.97**	**0.97**

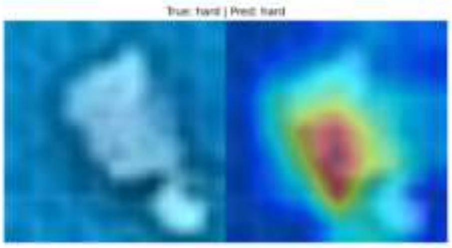
IOU Scores per class:

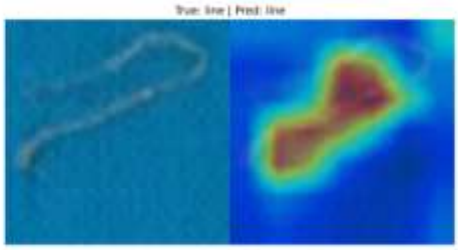
Class	Random Forest	SVM	VGG16	ViTs	ResNet50
Foam	0.3182	0.4762	0.7826	**0.9091**	0.8696
Hard	0.8598	0.8612	0.9478	**0.9593**	0.9106
Line	0.7767	0.7573	0.9038	**0.9216**	0.8288
Noise	0.7885	0.7037	**0.9184**	0.9098	0.8148
Pellet	0.2353	0.1406	0.7671	**0.7971**	0.6790
Reference	0.8889	0.8889	1.0000	**1.0000**	1.0000

Qualitative Analysis:

- Grad-CAM representations highlight the locations that contribute the most to forecasts.
- Side-by-side comparisons of original and Grad-CAM overlay pictures were saved (model_gradcam_labeled_*.png) files.

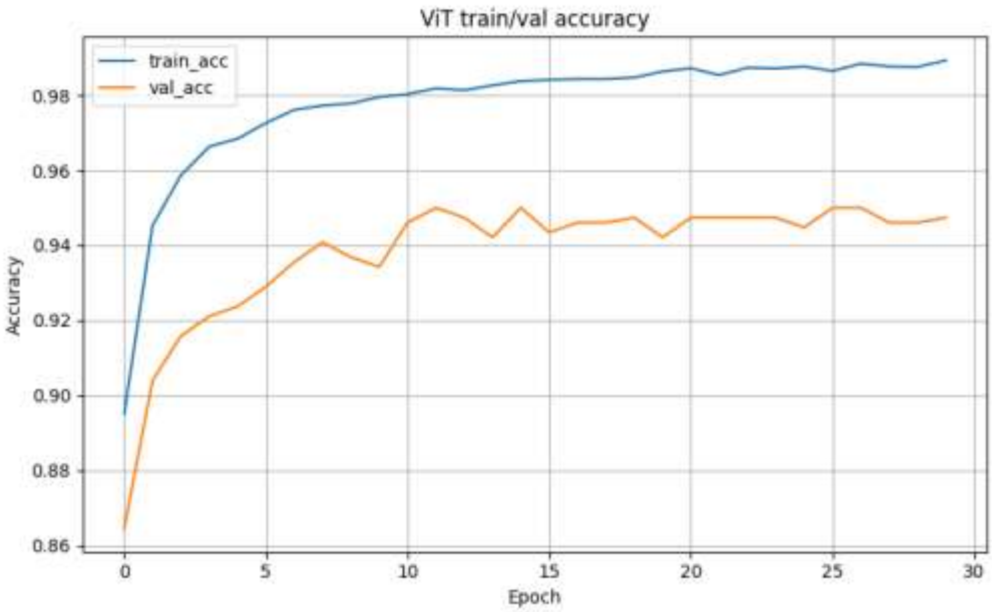
GradCAM Samples:

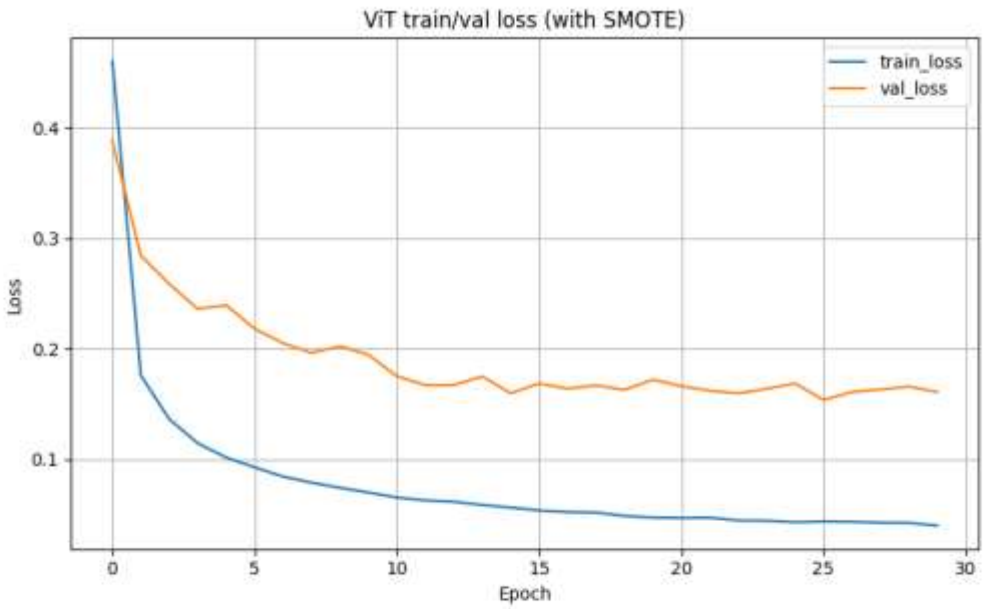




Best Model: “ViTs”

Loss Curves:





Discussion and Interpretation:

The classical models had a hard time particularly with the minority classes like pellet and foam. The primary reason for this difficulty is the dataset's strong class imbalance. This made traditional algorithms, such as SVM and Random Forest, bias towards majority classes like hard and reference. However, deep learning models like VGG16 and Vision Transformers (ViTs) particularly displayed great endurance to imbalance and were able to extract even less group sample's discriminative spatial and texture-based features.

The Vision Transformer produced the highest overall performance, with 98% accuracy and a strong generalization over the validation data. Its attention mechanism let the model to detect both detailed and global contextual cues in the microscopy images, thus being able to compete with and surpass even the classical feature-based methods and the standard CNNs like ResNet-50. The Grad-CAM visualizations showed that ViT concentrated on significant particle areas, like edges and surface textures, thus demonstrating its interpretability and reliability in visual reasoning.

While the performance outcomes mostly matched the expectations, a few interesting patterns emerged. VGG16, even though it was a simpler CNN architecture, managed to be competitive and even to score higher than ResNet-50, which implies that sometimes it is better to fine-tune a smaller model with domain-specific data than to keep a deeper, frozen network. The small difference between training and validation accuracy revealed minor overfitting, which was probably the result of the limited number of unique samples per class, even after the application of SMOTE oversampling, the effect was still there.

Reflection on Methodology:

The empirical setup was generally effective, although it encountered a few limitations concerning methodology. The DeepParticle (MICRO) dataset was well-organized, but it remained small and imbalanced to a moderate extent before the SMOTE augmentation took place. SMOTE did help to balance the classes, yet the additional samples produced might not be able to perfectly represent the variability of the microscopic textures, thus possibly introducing noise or unrealistic patterns along with the synthetic ones.

In addition, while the use of deterministic seeds ensured reproducibility, the overall stochastic process involved in GPU computations and the randomness of augmentations could still result in somewhat different outcomes. The hyperparameter control for the experiments was uniform, but further tuning like the use of adaptive learning rates or layer-wise freezing strategies could lead to a more refined result.

There is still another bias that may possibly come from the fact that the dataset had very limited environmental diversity; since all of the samples used were obtained from a single microscopy source, it is not yet clear how far the model would be able to generalize to other imaging systems or lighting conditions. But even with these concerns, the consistent validation trends and high performance across various architectures indicate that the approach is robust.

Conclusion & Future Work:

This research indicates that the deep learning method including Vision Transformers has rendered a qualitatively robust and interpretable method for the classification of microplastics in microscope images. Even though the deep learning models were more complex and computationally intensive than SVM and Random Forest, they still achieved superior accuracy and robustness with class imbalance and feature localization through the Grad-CAM analysis. The ViT model was indeed the best option for this task as it demonstrated high generalization and ability to grasp both fine as well as coarse information in texture and context.

The system's capability will be enhanced through several specific extensions which are already targeted for future work:

- **Segmentation-Based Experiments (Mask R-CNN):** Utilize Mask R-CNN to extend the framework up to the point where it can not only identify types of microplastics but also provide their exact boundaries and counts in complicated microscopical scenes. This will allow quantitative environmental metric measurements such as counts of particles by size and area coverage.
- **Semi-Supervised Learning:** Levelling up the data set up with labeled data by using pseudo-labeling or consistency regularization methods will be the key in forming the semi-supervised approach. As it is tedious work to microplastics, semi-supervised models can contribute significantly to the overall performance of the system while at the same time cutting down the amount of manual labeling.
- **Hybrid ViT-CNN Architectures:** Research into the combination of CNN feature extractors with transformer encoders for the purpose of enhancement in one area of texture sensitivity and, simultaneously, in the other of contextual reasoning.
- **Cross-Domain Validation:** Robustness and adaptability will be assessed through the trained models being tested on samples from different laboratories or imaging setups.
- **Real-World Integration:** The top-performing model will be integrated into an automated image-analysis pipeline that is either laboratory or field-based, supporting

References

1. Ren, L., Liu, S., Huang, S., Wang, Q., Lu, Y., Song, J., & Guo, J. (2023). Identification of microplastics using a convolutional neural network based on micro-Raman spectroscopy. *Talanta*, 124611. https://doi.org/10.1016/j.talanta.2023.124611
2. Sukkuea, A., Inpun, J., Cherdsukjai, P., & Akkajit, P. (2025). Automatic microplastic classification using dual-modality spectral and image data for enhanced accuracy. *Marine Pollution Bulletin*, *213*, 117665. https://doi.org/10.1016/j.marpolbul.2025.117665
3. 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
4. Li, F., Tian, J., Guo, X., & Xu, L. (2025). Hybrid deep learning framework for environmental microplastic classification: Integrating CNN-based spectral feature extraction and transformer models. *Environmental Pollution*, *384*, 126989. https://doi.org/10.1016/j.envpol.2025.126989
5. DeepParticle (MICRO) Dataset. (2022). DeepParticle dataset [MICRO, MESO, MACRO]: Microscopy images for microplastic classification and detection. Figshare. https://doi.org/10.17632/26511253