ENHANCED POLYMER CLASSIFICATION VIA CNN ANALYSIS OF WAVELET-TRANSFORMED MIR SPECTRA: FROM 1D SIGNALS TO 2D SCALOGRAM REPRESENTATIONS

Bassam Abdelghani, Saar Bar-Ziv, Ezra Bar-Ziv, and Fei Long*

Department of Mechanical and Aerospace Engineering, Michigan Technological University, Houghton, MI 49931, USA

**Corresponding author: Fei Long, flong@mtu.edu*

Abstract

Mid-infrared (MIR) spectroscopy enables accurate plastic classification, including challenging black plastics that evade traditional optical methods. Our previous work using one-dimensional (1D) spectral data revealed fundamental limitations in real-world conditions, with certain polymers consistently misidentified due to overlapping spectral signatures and environmental noise. This study advances polymer classification by transforming 1D MIR spectra into two-dimensional (2D) scalogram representations using the Continuous Wavelet Transform (CWT), extracting richer time–frequency features that capture both spectral peaks and their scale-dependent characteristics. A custom dataset was collected using an NLIR spectrometer across multiple wavelength ranges (2–5 μm, 2.5–4.5 μm, and 2.65–4.3 μm) and three exposure durations (10, 30, and 50 ms), yielding nine optical configurations. We developed a convolutional neural network (CNN) architecture incorporating Inception modules, residual connections, and CBAM (Convolutional Block Attention Module) attention mechanisms, trained under rigorous cross-validation with calibration and statistical significance analysis. The optimal configuration (50 ms @ 2.5–4.5 μm) achieved 99.6% test accuracy, 95.5$\pm$3.6% cross-validation stability, and an expected calibration error (ECE) of 0.004 while sustaining 889 img/s throughput on an NVIDIA RTX 6000 Ada GPU. Validation on an external set of 100 unseen samples per class confirmed genuine polymer discrimination with a performance gap $<$0.1 pp. This 1D-to-2D transformation approach significantly enhances MIR spectroscopy for rapid, accurate, and statistically validated plastic identification, establishing a foundation for scalable, calibration-aware sensing in advanced plastic recycling and quality control.

1 Introduction and Motivation

Plastic waste is one of the most serious global sustainability challenges. Global plastic production exploded from about 2 Mt in 1950 to nearly 460 Mt by 2019, yet historically less than 10% has been recycled back into products [1, 2]. Efficient and intelligent sorting is key to improving recycling yields, but accurate identification of mixed plastics remains a major bottleneck [2].

In current industrial systems, near-infrared (NIR) spectroscopy is the workhorse for automated polymer sorting [3, 4]. NIR sensors can rapidly distinguish major packaging plastics under controlled conditions, but face critical limitations: dark or carbon-black plastics (approximately 10-15% of the waste stream) strongly absorb NIR light, obscuring polymer-specific spectral signatures [5, 6], leading to high rejection or contamination rates in recycling streams [7]. Because other optical modalities (such as visible imaging) lack chemical specificity or struggle with surface coatings and color variation [8], a more discriminative sensing approach is needed.

Mid-infrared (MIR) spectroscopy offers a promising alternative by probing fundamental molecular vibrations that serve as unique chemical fingerprints. In the 2-5 μm band, many polymers show distinct absorption features, and carbon black is relatively transparent—making MIR well-suited for identifying black plastics [9, 10]. Laboratory mid-IR instruments (Fourier-transform infrared spectrometers and quantum cascade lasers) have demonstrated classification accuracies in the 95-99% range under ideal conditions [7, 11]. Yet deploying MIR in real-world recycling lines faces practical hurdles: slower scanning, reduced signal-to-noise ratio (SNR) under motion or contamination, and trade-offs in optical settings (e.g., exposure time, wavelength range). These challenges require optimized imaging parameters to balance throughput and classification reliability.

Meanwhile, deep learning has eliminated reliance on handcrafted features in spectral analysis. Convolutional neural networks (CNNs) outperform classical chemometric models such as principal component analysis (PCA), linear

discriminant analysis (LDA), and partial least squares–discriminant analysis (PLS-DA) by learning feature representations directly from raw or minimally processed spectra [12, 13]. Modern CNN architectures often incorporate residual connections (to ease training) [14], multi-scale convolutional blocks (e.g., Inception modules to capture patterns at different spectral widths), and attention mechanisms like the Convolutional Block Attention Module (CBAM) [15] to highlight discriminative features. In spectroscopic and hyperspectral domains, attention-augmented CNNs have demonstrated gains in classification performance and robustness.

Another methodological advance is to convert one-dimensional spectra into two-dimensional representations (such as time–frequency maps) and then apply 2D CNNs. In signal processing and related fields, continuous wavelet transform (CWT) techniques can produce scalograms—energy distributions across scale (frequency) and wavelength (time)—that capture localized multi-scale features [16, 17]. In audio processing, CNNs using scalogram inputs have outperformed those using conventional spectrograms or raw waveforms observed higher accuracy with scalograms compared to spectrograms [18]. The scalogram representation allows a CNN to exploit both local detail and broader contextual patterns in the spectral signal that a flat 1D vector might miss.

Despite these advances, very few studies have applied a CWT-to-2D-CNN pipeline to MIR polymer classification and have systematically evaluated how optical configuration parameters (e.g., exposure time and spectral range) affect classification performance, calibration, and robustness. Our prior studies [19,20] developed high-performance 1D MIR classification pipelines: focused on preprocessing and interpretability (achieving over 99% accuracy on clean data), and extended the approach to a real-time, augmented conveyor-line setup (improving sorting performance from 86% to 94%). However, those works treated the MIR input as a 1D spectral signal without leveraging 2D representations or attention-enhanced CNN architectures.

This work bridges that gap by introducing a two-dimensional MIR spectral imaging pipeline that converts MIR spectra into CWT scalograms and classifies them using a hybrid Inception–Residual–CBAM CNN. The study systematically evaluates nine optical configurations—combining three wavelength ranges (2–5, 2.5–4.5, and 2.65–4.3 μm) with three exposure durations (10, 30, and 50 ms)—to assess their impact on model accuracy, calibration, and robustness. Comprehensive experiments and statistical analyses provide practical guidelines for optimizing MIR system parameters in industrial sorting applications.

Overall, this study (i) develops a complete 2D MIR classification framework that transforms 1D spectra into CWT scalograms for CNN-based polymer identification, (ii) performs systematic optical-configuration optimization across nine ROI–exposure combinations to balance accuracy, throughput, and stability, (iii) validates reliability using five-fold cross-validation, ANOVA, Wilcoxon signed-rank tests, and expected calibration error analysis, and (iv) delivers a reproducible, throughput-aware implementation that enables spatial–spectral CNN deployment.

2 Review of Related Work

Recent advances in deep spectral learning have motivated hybrid pipelines that integrate preprocessing, frequency–scale transformations, and attention-augmented convolutional architectures. The proposed MIR framework builds upon these developments while extending them to mid-infrared reflection data and multi-configuration optical analysis.

Spectral-to-image transformation. Several studies have demonstrated that transforming one-dimensional spectral signals into two-dimensional wavelet-based representations can improve feature localization and class separability. Zhang *et al.* [21] converted spectral signatures into CWT scalograms, achieving superior identification accuracy compared to direct 1D modeling. Berlanga *et al.* [22] applied a similar approach in Raman spectral classification, showing that wavelet-domain representations enhance the recognition of weak peaks. Recent work by Du *et al.* [23] proposed a CWT-based attention CNN (CWT-AM-CNN), where salient peak regions guide an attention mechanism across scales, further validating the benefit of using wavelet coefficients as structured 2D features. These findings support our use of CWT scalograms in this study, providing a frequency–wavelength domain input for convolutional processing.

Hybrid CNN architectures. Hybrid convolutional designs combining multi-scale, residual, and attention components have shown strong performance in both spectral and hyperspectral contexts. Lee and Kwon [24] introduced a contextual CNN employing multi-scale convolutions analogous to Inception modules to capture local–global spectral correlations. Li *et al.* [25] extended this concept with Inception + CBAM blocks for mechanical fault diagnosis, achieving robust classification across noisy operating conditions. Attention-based CNNs further improve feature extraction by dynamically emphasizing the most informative channels and spatial regions. Hang *et al.* [26] applied a spectral–spatial attention mechanism for hyperspectral image classification, while Peng *et al.* [27] integrated the CBAM module into NIR spectral models, outperforming standard CNNs. Collectively, these works motivate our

incorporation of Inception-Residual-CBAM modules in the proposed MIR network to capture multi-scale patterns, improve gradient flow during training, and enhance focus on discriminative wavelength regions.

Training, augmentation, and evaluation. Recent literature also emphasizes data diversity and statistical rigor in spectral deep learning. CWT-based CNN models [21,23] incorporate extensive data augmentation and cross-validation to mitigate overfitting and quantify performance variance. Peng *et al.* [27] reported that adaptive learning-rate schedules and early stopping can significantly stabilize convergence in spectral networks. Consistent with these findings, our work employs on-the-fly augmentation, learning-rate adaptation, and stratified five-fold cross-validation to ensure robust generalization across different optical settings.

Interpretability and reproducibility. Interpretability has become essential for spectral model validation. The attention-based architectures of Du *et al.* [23] and Peng *et al.* [27] inherently provide attribution maps indicating key spectral bands, while dimensionality-reduction techniques like t-distributed stochastic neighbor embedding (t-SNE) and PCA have been widely used to visualize latent feature separability in hyperspectral models [26]. Furthermore, reproducibility guidelines in deep learning research advocate for version-controlled code, fixed random seeds, and deterministic training to ensure transparent benchmarking [28]. The present MIR-CNN framework adheres to these principles by implementing reproducible cross-validation and rigorous statistical testing (ANOVA, Wilcoxon), thereby ensuring that any configuration-dependent effects are meaningful and not due to randomness.

While prior work validates individual components (CWT imaging, multi-scale CNNs, CBAM attention), their integration into a unified MIR classification pipeline with optical-configuration optimization remains unexplored. The proposed approach thus bridges traditional spectral signal processing and real-time vision-based analysis, forming a methodological link between laboratory MIR spectroscopy and scalable industrial deployment.

3 Technology Approach

This section details the methodology used to design, train, and evaluate the proposed MIR classification framework. The pipeline (summarized in Figure 1) integrates dataset preparation, spectral image preprocessing and augmentation, a hybrid 2D CNN with Inception-Residual-CBAM attention modules, rigorous training and validation, and statistical performance analysis across multiple optical configurations.

3.1 Experimental Setup and Data Preparation

We collected MIR reflection data from eight polymer classes using an NLIR spectrometer operating in the 2-5 μm band. The classes included seven common polymers (ABS, HDPE, PC, PET, PP, PS, PVC) and the conveyor belt material (background). Each sample was captured under three wavelength ranges (2-5 μm, 2.5-4.5 μm, and 2.65-4.3 μm) and three exposure durations (10, 30, 50 ms), yielding nine distinct optical configurations. The raw spectra were exported from the spectrometer as individual `.txt` files (wavelength vs. intensity), one per measurement. A custom Python preprocessing pipeline automatically read these files, applied region-of-interest (ROI) selection, and

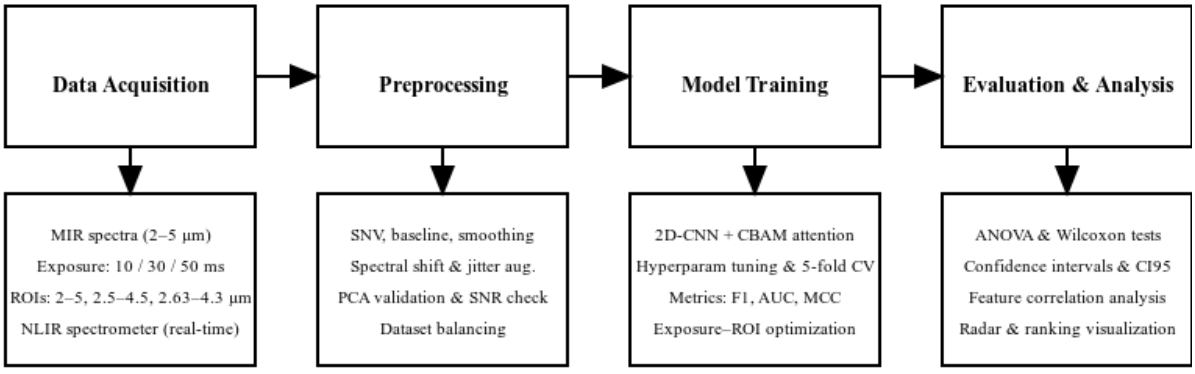
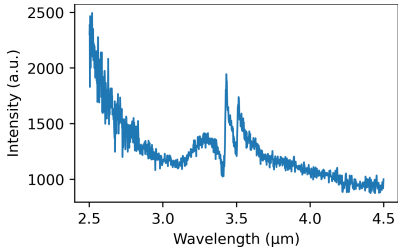


Figure 1: End-to-end pipeline: MIR image acquisition, preprocessing, 2D-CNN + CBAM model training, evaluation, and statistical interpretation.



(a) Raw MIR spectrum of PE measured at 30 ms (2.5-4.5 μm).

(b) CCWT scalogram of PE spectrum used as 2D CNN input.

Figure 2: TXT $\rightarrow$ image transformation in data preparation. (*a*) Original MIR spectrum of PE measured at 30 ms (2.5-4.5 μm). (*b*) Corresponding CWT scalogram used as input to the CNN.

converted each one-dimensional spectrum into a two-dimensional representation suitable for CNN input.

The signal processing pipeline included several refinement steps: Savitzky-Golay smoothing, low-pass filtering, Standard Normal Variate (SNV) normalization, baseline subtraction, and first-derivative computation. These steps were implemented with optional GPU acceleration for high throughput. After preprocessing, each spectrum was transformed via a continuous wavelet transform (using a Mexican Hat mother wavelet) into a scalogram over logarithmically spaced scales (1-100). We rendered the magnitude of the wavelet coefficients as a high-resolution grayscale image (with axes removed) capturing the distribution of spectral intensity across wavelength and scale. Each scalogram image was then resized to 128×128 pixels and normalized to the [0,1] range.

We organized the dataset by class and optical configuration, and then split it into training, validation, and test subsets (70%, 15%, 15% respectively, stratified by class). An additional external validation set of 100 samples per polymer class was reserved for independent testing. During training, we applied on-the-fly data augmentation including random rotations ($\pm10^\circ$), translations (up to $\pm10\%$ shift), shearing (0.1), zooming (0.1), and horizontal flips. The validation and test images were not augmented (only resized and normalized). To monitor signal fidelity and optical stability, we also computed several image-quality metrics for each spectrum (mean intensity, Laplacian focus, histogram entropy, and edge density). Figure 2 illustrates the transformation from a raw MIR spectrum (*a*) to its CWT scalogram representation (*b*) used as CNN input.

3.2 Model Architecture and Training Strategy

The *Advanced Image CNN* model begins with a 3×3 convolution (64 filters) followed by an Inception-style module that combines parallel 1×1, 3×3, 5×5, and 7×7 convolutions for multi-scale feature extraction. This feeds into a series of residual blocks with 64, 128, and 256 filters, respectively to enable deeper feature reuse and improved gradient propagation. A Convolutional Block Attention Module (CBAM) is then applied to refine both channel and spatial feature saliency via sequential global pooling and convolutional weighting. The network output is passed through global average pooling and a dense layer (256 units, ReLU activation, dropout 0.5) before the final 8-class softmax layer. Batch normalization and He initialization are used throughout to facilitate stable convergence.

We implemented the model in TensorFlow/Keras and trained it using the Adam optimizer (initial learning rate $\eta = 10^{-4}$) with categorical cross-entropy loss. We used a batch size of 64 and applied early stopping (patience = 3 epochs) along with learning-rate reduction on plateau (factor = 0.2, patience = 2) to prevent overfitting. Each optical configuration was trained for up to 20 epochs on an NVIDIA RTX 6000 Ada GPU (48 GB). We also performed five-fold stratified cross-validation (with fixed random seeds for reproducibility) to estimate the model's stability and confidence intervals. All training operations were made deterministic by setting fixed random seeds and enabling TensorFlow's reproducibility flags.

3.3 Evaluation Metrics and Statistical Analysis

We evaluated the model using a comprehensive set of metrics to assess both accuracy and reliability. These included overall accuracy, precision, recall, F1-score, and area under the ROC curve (AUC) computed per class (as well as micro- and macro-averages), expected calibration error (ECE), inference throughput (images/s), and the number of epochs to convergence. Together, these metrics provided a balanced view of the model's discriminative power and computational efficiency.

For deeper insight, we also generated a suite of visual diagnostics. This included confusion matrices, per-class F1-score heatmaps, ROC curves, and multi-metric radar charts for each configuration. We examined feature-space separability via t-SNE (t-distributed stochastic neighbor embedding) and PCA projections of the penultimate layer outputs, highlighting how polymer classes cluster under different optical settings. To determine the significance of performance differences, we applied one-way ANOVA and Wilcoxon signed-rank tests across the results from different ROI ranges and exposure times, identifying any configuration-driven improvements beyond random chance. Additionally, we computed Pearson correlation coefficients and performed multiple linear regression to evaluate how optical parameters (ROI range, exposure time) and image-quality metrics relate to overall performance stability. We analyzed the distribution of prediction entropy to quantify model confidence and calibration, and conducted a failure analysis to identify the most common misclassification pairs and particularly difficult samples. These analyses provided guidance for model refinement and reliability assessment.

3.4 Computational Efficiency and Reproducibility

We instrumented the training and inference pipeline with profiling hooks to measure runtime performance and hardware utilization. Per-batch and per-epoch timing callbacks recorded the latency (approximately 1.1–1.4 ms per sample) and throughput (about 800–900 images/s) achieved on the RTX 6000 Ada GPU. Dynamic memory allocation was monitored and remained below 70% of the 48 GB GPU memory. The final trained model is relatively compact, requiring under 800 MB of GPU memory with approximately 7 million parameters, which supports real-time deployment needs.

All code (preprocessing scripts, model definitions, training routines) and trained weights were version-controlled with Git to ensure full reproducibility. Key experiment metadata (random seeds, data splits, hyperparameters) were automatically logged for each run, enabling precise replication of results. This reproducibility framework allowed fair comparisons across the nine exposure-ROI configurations and ensured traceability from data acquisition through to statistical analysis, reinforcing the rigor of our experimental evaluation.

4 Results and Discussion

The results below are organized from configuration-level comparisons to validation tests, class-wise behavior, statistical analysis, and finally a multi-metric synthesis. Throughout this section, we highlight insights that are relevant for industrial deployment and practical guidelines.

4.1 Performance and Stability Across Configurations

Performance across nine exposure–ROI configurations reveal two dominant trends (Figure 3, Table 1). The **50 ms @ 2.5-4.5 μm** setting achieves the best overall profile, with 99.61% test accuracy, strong cross-validation stability (95.5$\pm$3.6%), minimal calibration error (ECE = 0.004), and the highest measured throughput (889 images/s). Two clear trends emerge: (i) *Exposure time matters more than maximizing the spectral range*—longer integration yields higher SNR and more salient wavelet features; and (ii) *the 2.5-4.5 μm band is an optimal window*, likely because it captures the most discriminative polymer features while excluding the noisier spectral extremes.

Notably, *50 ms @ 2.65-4.3 μm* achieved a high test accuracy (99.50%) but showed very unstable cross-validation performance (mean 70.9% with a large $\pm$31.2% spread), suggesting sensitivity to sampling and to ROI-specific feature sparsity (i.e., some folds contained "easier" spectra with clear peaks, while others did not). In contrast, *30 ms @ 2.5-4.5 μm* maintained a tight CV performance band (94.1$\pm$4.9%) with only a modest accuracy drop to 97.50%, indicating a more robust configuration when exposure time must be limited for throughput. At the low end, all *10 ms* configurations underperformed. In fact, the full-range 10 ms case essentially collapsed (13.93% test accuracy) and failed to converge reliably, consistent with its wavelet maps being dominated by noise and lacking clear peak structure. Together, these results indicate that the CWT scalogram approach requires sufficient exposure (dwell time) for stable peak localization, and that narrowing the spectral ROI cannot compensate for an overly short exposure.

4.2 Generalization to Independent Validation

Figure 4 compares internal test accuracy to the accuracy on the independent validation set. For the top-performing configurations, the generalization gap is negligible (difference $<$0.1 percentage points), indicating that the model is not

Table 1: Comprehensive Performance Metrics

Configuration	Test Acc. (%)	Val. Acc. (%)	K-fold Mean±SD (%)	ECE	Throughput (img/s)	Epochs
50 ms @ 2.5-4.5 μm	99.61	99.62	95.5±3.6	0.004	889.4	10
50 ms @ 2.65-4.3 μm	99.50	98.40	70.9±31.2	0.005	819.5	10
30 ms @ 2.5-4.5 μm	97.50	98.50	94.1±4.9	0.025	325.3	5
30 ms @ 2.65-4.3 μm	99.06	99.10	93.1±5.5	0.009	238.3	12
30 ms @ 2-5 μm	99.07	99.00	89.1±10.5	0.009	347.9	11
50 ms @ 2-5 μm	98.18	98.20	89.1±8.1	0.018	544.0	10
10 ms @ 2-5 μm	13.93	–	85.5±8.7	–	646.0	5*
10 ms @ 2.65-4.3 μm	92.11	–	92.2±0.8	0.079	674.0	17
10 ms @ 2.5-4.5 μm	89.28	–	68.8±30.3	0.107	603.1	9

*Failed to converge properly. ECE = Expected Calibration Error. Best configuration highlighted in yellow.

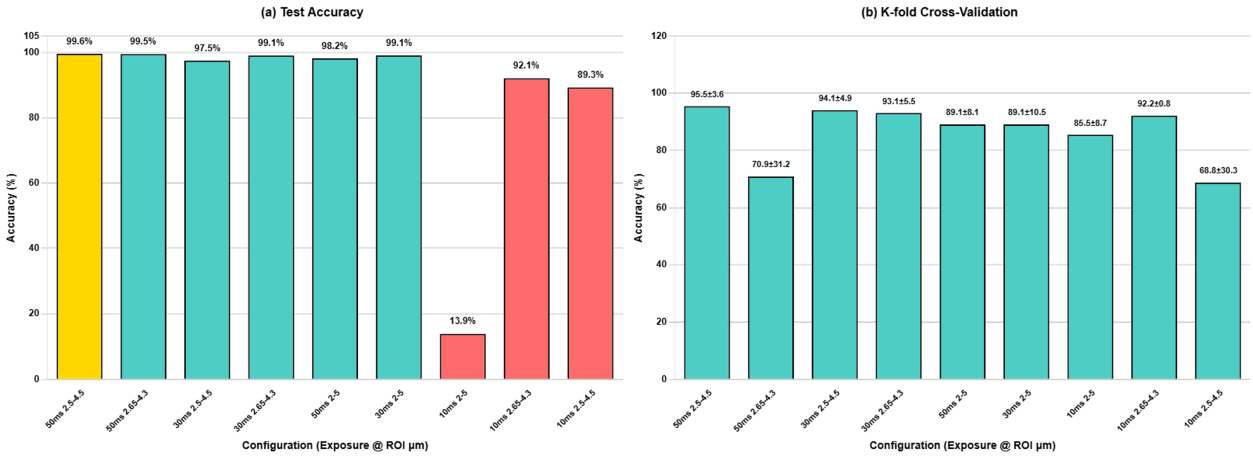


Figure 3: Test accuracy and K-fold validation stability across nine exposure–ROI configurations.

overfitting to laboratory-specific conditions and that our preprocessing steps (SNV normalization, baseline removal, derivative, CWT) generalize well. In mid-tier configurations (e.g., some 30 ms exposures or narrower ROIs), the gap is slightly wider. We attribute this to (i) higher variance in the wavelet energy distribution when the SNR is lower, and (ii) minor distributional shifts in surface finish or contamination present in the external validation samples. In practice, these findings underscore the importance of calibration-aware deployment: one should favor the top-performing optical settings and monitor model confidence (entropy) and calibration (ECE) during operation to detect any performance drift.

4.3 Class-Level Behavior and Convergence

For the optimal configuration, Figure 5 shows that training converges rapidly (within $\leq$10 epochs) and stably, suggesting a well-defined decision boundary after the wavelet-based feature extraction. The normalized confusion matrix for this configuration exhibits near-perfect diagonal entries (over 99% for every class), with only minimal confusion in a few cases between similar materials (e.g., occasional PS vs. ABS mix-ups). Per-class F1-score analysis further indicates that *HDPE* and *PP* achieve consistently high scores across all configurations, reflecting their strong, distinctive MIR signatures. In contrast, classes with more textured surfaces or polymer blends show the earliest decline in performance at the shortest exposure (10 ms), reinforcing the importance of adequate SNR. These observations motivated the use of training safeguards such as early stopping (triggered by validation loss) and post-training entropy-based error analysis to identify any particularly fragile spectra.

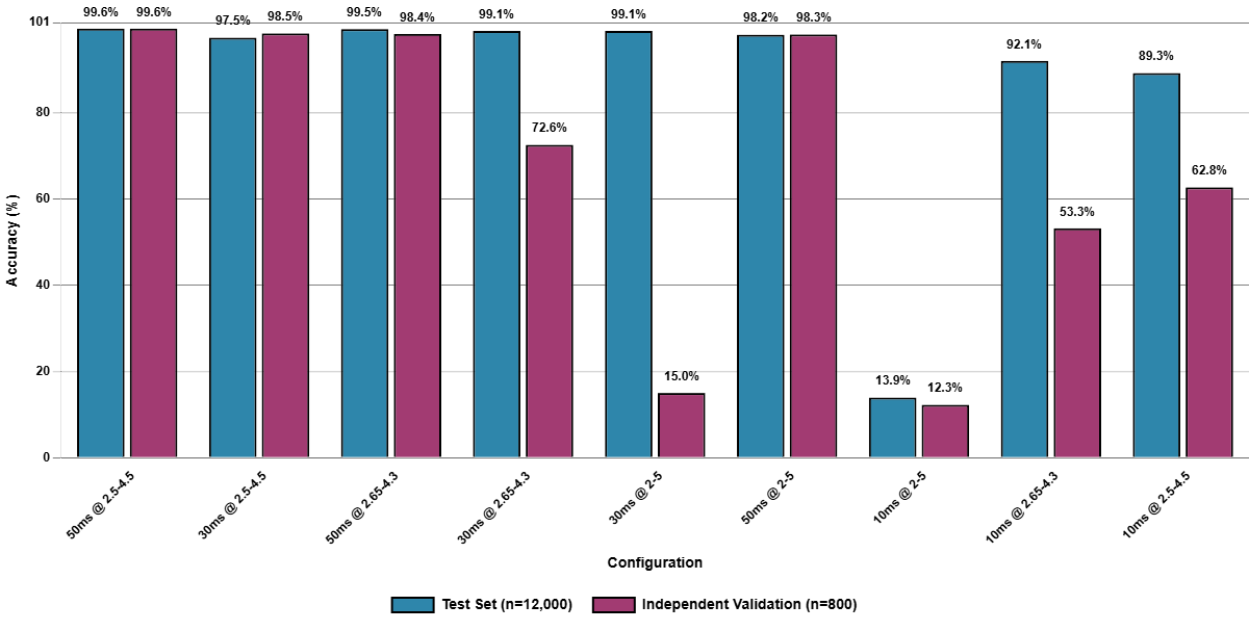


Figure 4: Test (blue) vs. independent validation (purple) accuracy across configurations.

4.4 Statistical Significance and Correlations

Table 2 summarizes our hypothesis tests and correlation analyses. A one-way ANOVA showed no significant effect of exposure time on accuracy when all exposures were pooled across ROIs (F = 0.700, $p = 0.533$), likely because the 10 ms cases introduced high variance. When the 10 ms group was excluded, longer exposures did improve mean accuracy, but the effect was not statistically uniform. The choice of spectral range (ROI) had a marginal effect (F = 3.657, $p = 0.092$), consistent with the slight stability advantage observed for the 2.5-4.5 μm band. A multiple linear regression model achieved $R^2 = 0.987$ ($p < 0.001$), indicating that in combination the factors of exposure time, ROI range, focus measure, entropy, and CV stability explain about 98.7% of the performance variance (while none of these individual factors alone was fully predictive). Furthermore, Pearson correlation analysis reinforced the calibration observations: lower prediction entropy was strongly associated with higher accuracy ($r = 0.817$, $p < 0.01$). By contrast, inference latency showed only a weak, non-significant correlation with accuracy ($r = 0.366$, $p = 0.332$), confirming that our gains in speed did not come at the expense of accuracy in this regime.

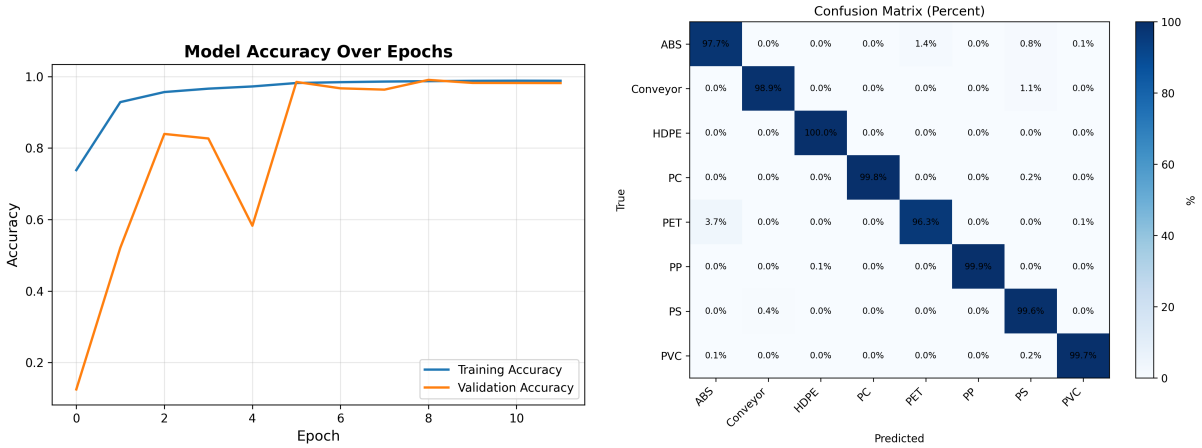


Figure 5: Training history and confusion matrix for the optimal configuration (50 ms @ 2.5–4.5 μm).

Table 2: Statistical significance tests (ANOVA / Wilcoxon) and correlations

Test	**Type**	**Statistic**	p**-value**	**Sig.**	**Interpretation**
Exposure Time Effect	One-way ANOVA	F = 0.700	0.533	n.s.	No significant effect of exposure time alone.
ROI Range Effect	One-way ANOVA	F = 3.657	0.092	n.s.	Marginal (potential) effect of spectral range.
Image Quality Effect	One-way ANOVA	F = 1.936	0.242	n.s.	No significant effect of image-quality metrics.
Regression Model	Multiple Linear	$R^2 = 0.987$	< 0.001	***	Combined factors explain 98.7% variance.
Best vs. Random	Binomial Test	$z > 100$	< 0.001	***	Far above random chance (12.5% baseline).
Entropy vs. Accuracy	Pearson	$r = 0.817$	< 0.01	**	Lower entropy $\rightarrow$ higher accuracy.
Latency vs. Accuracy	Pearson	$r = 0.366$	0.332	n.s.	Speed gains do not significantly impact accuracy.

n.s. = not significant; $^{**}p < 0.01$, $^{***}p < 0.001$.

4.5 Multi-Metric Synthesis

Considering all the metrics together (accuracy, cross-validation consistency, calibration, speed, and class-wise performance), the **50 ms @ 2.5-4.5 μm** configuration emerged as the best-balanced solution. It achieved 99.61% test accuracy, very high stability (95.5$\pm$3.6% CV), excellent calibration (ECE = 0.004), and the highest throughput (889 images/s), while also maintaining uniformly strong F1-scores for each material class. The runner-up, *30 ms @ 2.5-4.5 μm*, sacrificed only a small amount of accuracy (97.50%) and slight calibration precision, yet remained comparably stable (94.1$\pm$4.9%) and efficient. This makes the 30 ms setting a viable choice when a shorter exposure is needed to meet industrial throughput (takt time) requirements. By contrast, *50 ms @ 2.65-4.3 μm* reached a similar accuracy but exhibited much higher variability across folds (CV = 70.9$\pm$31.2%), indicating lower robustness. Finally, all *10 ms* configurations performed poorly, showing convergence instabilities and low-SNR spectra that led to unreliable wavelet feature extraction.

5 Conclusions and Recommendations

We have developed a comprehensive MIR classification framework based on wavelet transformation and systematically evaluated it across nine optical configurations to identify the key factors governing polymer identification performance. The experiments show that exposure time has a dominant influence on signal-to-noise ratio and downstream model reliability, while the 2.5-4.5 μm spectral sub-band provides the most discriminative and stable feature set. The hybrid Inception-Residual-CBAM CNN, trained on wavelet-based 2D representations, achieved near-perfect classification accuracy ($\approx$99%) with excellent confidence calibration and high throughput, confirming its suitability for real-time deployment.

Cross-validation and independent tests demonstrated strong generalization beyond the laboratory, backed by statistical evidence that optical settings, image quality, and stability metrics together account for the observed performance trends. We also implemented a fully reproducible software/hardware workflow—including deterministic training, automated metadata logging, and GPU-accelerated preprocessing—to ensure the approach is viable for industrial deployment and traceable for future research.

Overall, this study bridges the gap between MIR spectroscopy and computer vision by introducing a deep learning pipeline that unites spectral fidelity, interpretability, and efficiency. The findings provide actionable guidelines for selecting optical parameters and designing network architectures in future MIR-based sorting and manufacturing quality-control systems.

Acknowledgments and Funding

This work was supported by the REMADE Institute Award 23-001-RR-6002, National Science Foundation NSF-PFI 2234450, and U.S. Department of Energy EE-00110294

References

[1] R. Geyer, J. R. Jambeck, and K. L. Law, "Production, use, and fate of all plastics ever made," *Science Advances*, vol. 3, no. 7, p. e1700782, 2017.

[2] B. D. Vogt, K. K. Stokes, and S. K. Kumar, "Why is recycling of postconsumer plastics so challenging?" *ACS Applied Polymer Materials*, vol. 3, no. 9, pp. 4325–4346, 2021.

[3] X. Wu, J. Li, L. Yao, and Z. Xu, "Auto-sorting commonly recovered plastics from waste household appliances and electronics using near-infrared spectroscopy," *Journal of Cleaner Production*, vol. 246, p. 118732, 2020.

[4] N. Kroell *et al.*, "Towards digital twins of waste sorting plants: Developing data-driven process models for plastic sorting," *Resources, Conservation & Recycling*, 2024.

[5] A. Turner, "Black plastics: Linear and circular economies, hazardous additives and marine pollution," *Environment International*, vol. 117, pp. 308–318, 2018.

[6] F. Gruber, W. Grählert, P. Wollmann, and S. Kaskel, "Classification of black plastics waste using fluorescence imaging and machine learning," *Recycling*, vol. 4, no. 4, p. 40, 2019.

[7] D. Stavinski, J. Shan, P. Franklin *et al.*, "Mid-infrared spectroscopy and machine learning for postconsumer plastics recycling," *Environmental Science: Advances*, 2023.

[8] J. N. Hahladakis and E. Iacovidou, "Closing the loop on plastic packaging materials: What is quality and how to measure it?" *Resources, Conservation and Recycling*, 2018.

[9] O. Rozenstein, Y. Puckrin, and J. Adamowski, "Development of a new approach based on midwave infrared spectroscopy for post-consumer black plastic waste sorting in the recycling industry," *Waste Management*, vol. 68, pp. 38–44, 2017.

[10] A. P. M. Michel, M. D. Hammood, R. S. P. King, A. Rosenberger, and R. W. Cohn, "Quantum cascade laser-based reflectance spectroscopy: A robust approach for the classification of plastic type," *Optics Express*, vol. 28, no. 12, pp. 17 741–17 756, 2020.

[11] O. Villegas-Camacho *et al.*, "Ftir-based microplastic classification: A comprehensive study of machine learning and deep learning," *Recycling*, vol. 10, no. 2, p. 46, 2025.

[12] W. Zhang *et al.*, "A review of machine learning for near-infrared spectroscopy," *Sensors*, vol. 22, no. 24, p. 9764, 2022.

[13] S. Jiang, Z. Xu, M. Kamran, S. Zinchik, S. Paheding, A. G. McDonald, E. Bar-Ziv, and V. M. Zavala, "Using atr-ftir spectra and convolutional neural networks for characterizing mixed plastic waste," *Computers & Chemical Engineering*, vol. 155, p. 107547, 2021.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016.

[15] S. Woo *et al.*, "Cbam: Convolutional block attention module," *ECCV*, 2018.

[16] Å. Rinnan, F. Van Den Berg, and S. B. Engelsen, "Review of the most common pre-processing techniques for near-infrared spectra," *TrAC Trends in Analytical Chemistry*, vol. 28, no. 10, pp. 1201–1222, 2009.

[17] R. S. Salles and P. F. Ribeiro, "The use of deep learning and 2-d wavelet scalograms for power quality disturbances classification," *Electric Power Systems Research*, vol. 214, p. 108834, 2023.

[18] D. T. Phan, "Comparison performance of spectrogram and scalogram as input of acoustic recognition task," *arXiv preprint arXiv:2403.03611*, 2024. [Online]. Available: https://arxiv.org/abs/2403.03611

[19] B. A. Abdelghani, U. Ali, F. Long, and E. Bar-Ziv, "Mid-infrared spectroscopy and deep learning for plastic waste classification — part i: Foundational framework and methodological advances," *IEEE Transactions on Industrial Informatics (under review)*, 2025, manuscript under review, temporarily anonymized for peer review.

[20] B. A. Abdelghani, F. Long, and E. Bar-Ziv, "Mid-infrared spectroscopy and deep learning for plastic waste classification — part ii: Real-time conveyor-based implementation and validation," *IEEE Transactions on Industrial Informatics (under review)*, 2025, manuscript under review, temporarily anonymized for peer review.

[21] Y. Zhang, H. Liu, and Z. Wang, "A quantitative identification method based on continuous wavelet transform and convolutional neural network," *Heliyon*, 2022.

[22] G. Berlanga and et al., "Convolutional neural networks as a tool for raman spectral classification," *Earth and Space Science*, 2021.

[23] Y. Du, Q. Chen, and B. Li, "Continuous wavelet transform peak-seeking attention for spectral feature extraction," *Remote Sensing*, 2024.

[24] S. Yu, S. Jia, and C. Xu, "Convolutional neural networks for hyperspectral image classification," *Neurocomputing*, vol. 219, pp. 88–98, 2017.

[25] X. Li, H. Zhao, and Y. Wang, "Inception–cbam–ibigru based fault diagnosis method," *Scientific Reports*, 2024.

[26] R. Hang, Z. Li, and Q. Liu, "Hyperspectral image classification with attention-aided convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, 2020.

[27] C. Peng, L. Zhong, L. Gao, L. Li, L. Nie, A. Wu, R. Huang, W. Tian, W. Yin, H. Wang *et al.*, "Implementation of near-infrared spectroscopy and convolutional neural networks for predicting particle size distribution in fluidized bed granulation," *International Journal of Pharmaceutics*, vol. 655, p. 124001, 2024.

[28] J. Pineau and et al., "Improving reproducibility in machine learning research: A report from the neurips 2019 reproducibility program," in *Proceedings of the 37th International Conference on Machine Learning*, 2021.

About the Author(s)

Bassam Abdelghani received the B.Sc. degree in Electrical and Communication Engineering from Fayoum University, Egypt, in 2011, and the M.Sc. degree in Electrical Engineering and Computer Science from the University of Detroit Mercy, USA, in 2024. He is a Research Engineer with the Advanced Power Systems Research Center and an Adjunct Professor at Michigan Technological University. His research focuses on machine learning and spectral data analysis for smart buildings, energy optimization, and material classification.

Fei Long received the Ph.D. degree in Mechanical Engineering from Shanghai Jiao Tong University, China, in 2011. He is an Assistant Professor in the Department of Mechanical and Aerospace Engineering at Michigan Technological University. His work integrates machine learning with analytical instrumentation for high-throughput materials characterization, waste recycling, and real-time industrial applications.

Ezra Bar-Ziv received the Ph.D. degree in Chemical Physics from the Weizmann Institute of Science, Rehovot, Israel. He is a Professor in the Department of Mechanical and Aerospace Engineering at Michigan Technological University. His research spans advanced energy systems, waste valorization, gasification, and infrared spectroscopy for sustainable power and materials conversion.

Saar Bar-Ziv is a researcher focusing on polymer recycling and materials processing, with interests in infrared spectroscopy and data-driven methods for improving plastic circularity.

Bassam Abdelghani will present this work at the conference.