Physics-Informed and Explainable Machine Learning for Mechanical Property Prediction in Recycled Polymer Composites

O. Tajir[1*], S. Elkoun[1], and T. Masrour[2,3]

[1] Center for Innovation in Technological Eco-Design (CITE), Department of Mechanical Engineering, Université de Sherbrooke, Sherbrooke, QC, Canada

[2] Laboratory of Mathematical Modeling, Simulation and Smart Systems (L2M3S), National Higher School of Arts and Crafts (ENSAM), University of Moulay Ismail, Meknes, Morocco

[3] Mathematics, Computer Science and Engineering Department, University of Quebec at Rimouski, Rimouski, QC, Canada

* *Corresponding author (oumayma.tajir@usherbrooke.ca)*

Keywords: *recycled polymer composites; physics-informed machine learning; feature engineering; explainable machine learning; formulation design*

1. Abstract

Mechanical recycling of polymer composites is challenged by significant variability in material properties arising from degradation, contamination, and heterogeneous formulations, which complicates formulation development and end-of-line qualification. As a result, optimization in recycled systems often relies on costly and time-consuming trial-and-error approaches. While machine learning (ML) provides powerful tools for property prediction, purely data-driven models often lack physical consistency and remain difficult to interpret. This work proposes a physics-informed and explainable ML framework for predicting the properties of recycled polymer composites. Physically motivated, property-aware descriptors derived from mixture theory are incorporated into the learning process, enabling a structured baseline-plus-correction modeling strategy. The framework is validated on industrial production data spanning diverse formulations, using batch-aware validation to account for correlated records and ensure realistic performance assessment. Compared to conventional composition-based models, the proposed approach improves predictive accuracy while remaining compact and interpretable. SHAP-based analysis further reveals physically consistent relationships between material characteristics and predicted properties, providing actionable insights for formulation design and supporting informed, data-driven decision-making in industrial contexts.

2. Introduction

Mechanical recycling of polymer composites is increasingly pursued to support material circularity [6]; however, recycled systems exhibit fundamentally different behavior from virgin materials. Degradation, contamination, and formulation heterogeneity introduce history-dependent and non-stationary property variations [7], complicating formulation development and industrial qualification. As a result, optimization still relies heavily on empirical, trial-and-error approaches [2]. Machine-learning techniques can model complex composition–property

relationships [3], but purely data-driven models often lack physical consistency and interpretability, limiting their reliability in industrial applications [8]. Physics-informed and explainable ML approaches address these limitations by explicitly incorporating domain knowledge into the learning process [9]. Unlike existing approaches, the proposed framework embeds physically-derived descriptors within a formulation-level representation, enabling improved generalization and intrinsic interpretability. Specifically, this work proposes (i) a physics-informed machine-learning framework for predicting mechanical properties of recycled composites, (ii) its validation on industrial data using batch-aware partitioning to avoid information leakage, and (iii) an interpretable analysis providing physically consistent insights into composition–property relationships.

3. Materials and Dataset

The dataset used in this study is derived from industrial production records from mechanical recycling operations of polymer composite materials. It includes common thermoplastic families such as polyolefins (PP, HDPE) and styrenics (PS, HIPS), representative of typical recycling streams. The data reflect real manufacturing conditions and the intrinsic variability of recycled systems arising from feedstock composition, processing history, and formulation complexity [7]. Each record includes rheological and mechanical indicators used for industrial qualification and formulation assessment [8]. As illustrated in Figure 1, the data are structured hierarchically: component-level descriptor vectors combine material, physicochemical, and processing information (Level 1), are extended to multi-component formulations (Level 2), and aggregated into formulation-level representations including composition fractions, additives, and process descriptors (Level 3). The dataset spans diverse formulations under varying conditions, with multiple records potentially originating from the same production batch. To prevent information leakage and ensure realistic evaluation, batch-aware data partitioning is applied during model development and testing [9]. The formulation-level representation is finally associated with target properties (Level 4), enabling prediction of mechanical and rheological responses. Standard preprocessing procedures are applied prior to modeling.

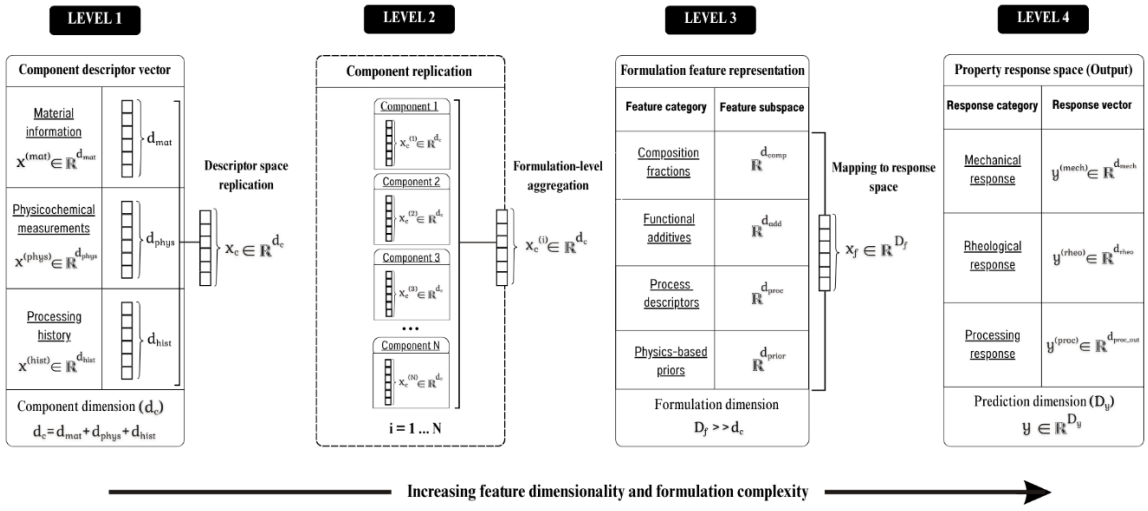

Figure 1. Hierarchical representation of recycled polymer formulation data.

4. Physics-Informed Feature Engineering

As illustrated in Figure 2, the proposed framework relies on physics-informed feature engineering to structure experimental data from recycled polymer systems prior to machine-learning modeling. Physics-guided approaches have been shown to improve both robustness and interpretability in materials modeling by reducing reliance on purely data-driven representations [10]. In this work, raw compositional variables are not used directly as model inputs. Instead, physically meaningful descriptors are constructed from experimental material data, including composition fractions (w_i), intrinsic material properties (P_i), and formulation-related parameters (φ_i), to capture the key mechanisms governing the behavior of recycled polymer composites. These descriptors include composition-weighted property averages ($\Sigma w_i P_i$), interaction terms capturing coupling effects between constituent phases ($w_i \cdot w_j$), and distribution-based measures such as heterogeneity indicators (H(w)). Together, these quantities encode composition effects, phase interactions, and formulation variability in a physically interpretable manner. The resulting descriptors are assembled into a structured input vector $X = [X_1, \ldots, X_i]^T$, which constrains the learning task to physically plausible relationships while remaining compatible with flexible nonlinear models. This representation shifts interpretability from a purely post hoc analysis to an intrinsic property of the input space, providing a transparent and physically grounded foundation for subsequent tree-based modeling and explainability analyses.

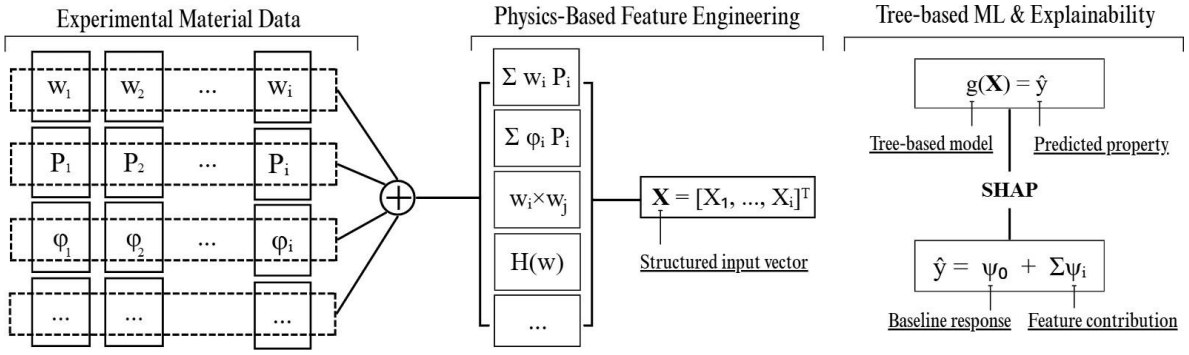

Figure 2. Overview of the proposed framework, illustrating the transformation of experimental material data into structured descriptors and their use for interpretable analysis of recycled polymer properties.

5. Model Performance and Evaluation

5.1. Model prediction performance

Model performance is assessed using a structured evaluation procedure combining data trimming, cross-validation, and independent testing, designed to reflect realistic industrial conditions. The dataset deliberately retains a degree of variability and noise, as such dispersion is intrinsic to recycled polymer manufacturing environments and should not be artificially removed. Trimming strategies are therefore applied in a controlled manner to limit the influence of extreme values while preserving representative industrial variability. For each trimming scenario, model hyperparameters are optimized through cross-validation on the training data, and final performance is evaluated on an independent test set. Tree-based ensemble methods are selected as reference models due to their robustness to nonlinear relationships and heterogeneous

tabular data, which are characteristic of recycled polymer systems. Based on test-set metrics, the three best-performing models are retained for each target property and summarized in Table 1 using consistent evaluation criteria. The analysis focuses on two representative properties, melt flow index (MFI) and flexural strength, selected to span distinct physical regimes, namely rheological and mechanical behavior. This choice enables evaluation of the proposed framework across fundamentally different material responses. The comparable performance observed across different learning algorithms highlights that predictive accuracy is primarily driven by the proposed physics-informed representation rather than by the choice of a specific model. Compared to composition-only inputs, the proposed physics-informed descriptors consistently improve predictive performance across all evaluated models.

Table 1. Test-set performance of the top-performing models for melt flow index and flexural strength.

Property	**Tree-based model**	**R^2 (test)**	**RMSE**	**MAE**
Melt Flow Index (MFI) (g/10 min)	Random Forest	0.85	5.64	3.72
	XGBoost	0.84	6.38	4.03
	ExtraTrees	0.83	5.99	3.90
Flexural Strength (FS) (Pa)	XGBoost	0.87	24,182	17,923
	Random Forest	0.86	24,401	18,772
	HistGradientBoosting	0.86	24,788	19,077

Figures 3a and 3b illustrate the agreement between predicted and experimental values for melt flow index (MFI) and flexural strength, respectively, using the best-performing model selected for each property. The observed dispersion primarily reflects intrinsic process and material variability rather than model instability, indicating that the proposed framework provides robust predictions under noisy, industrially representative conditions. This agreement is consistent with the R^2 values reported in Table 1.

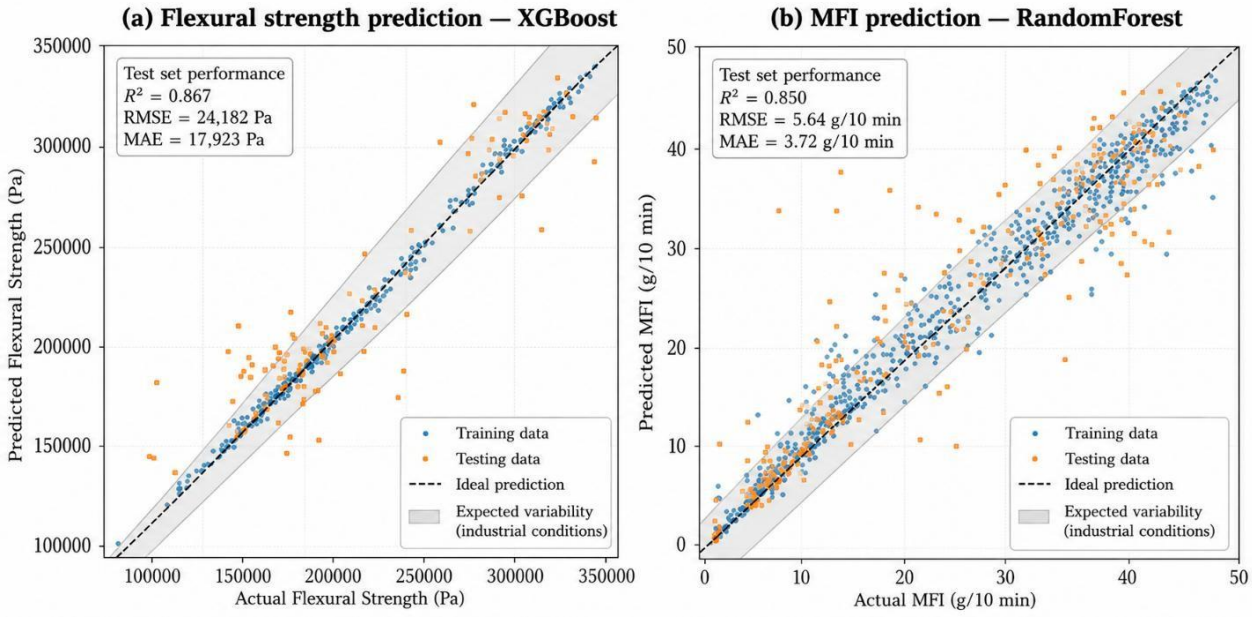

Figure 3. Parity plots for (a) flexural strength and (b) melt flow index (MFI), illustrating predictive performance under industrial conditions.

5.2. Physics-informed interpretation

Figures 4a and 4b provide a physics-informed interpretation of the proposed framework across two representative material properties spanning distinct physical regimes. In Figure 4a, flexural strength is expressed as a function of a physics-based descriptor derived from formulation bounds, with SHAP contributions capturing the combined influence of material structure and formulation interactions on mechanical performance [5]. The wide dispersion of SHAP contributions indicates that flexural strength is governed by coupled effects related to phase organization, material heterogeneity, and processing history, which collectively introduce significant variability in recycled polymer systems. In contrast, Figure 4b presents the melt flow index (MFI) as a function of a logarithmic prior descriptor representing the rheological state of the material. A clear monotonic relationship is observed, with SHAP contributions indicating that predictions are primarily driven by inherited flow characteristics. This behavior reflects the dominant role of molecular degradation history and viscosity evolution, resulting in a more direct and predictable response compared to mechanical properties. In both cases, the dashed curves represent smoothed trends of the model response, highlighting the underlying nonlinear structure of the data. Taken together, these results demonstrate that the proposed physics-informed representation enables interpretable predictions across fundamentally different material responses. While mechanical behavior is governed by complex structural interactions, rheological behavior remains largely controlled by prior material states. This distinction highlights the ability of the framework to capture different governing mechanisms while maintaining robustness under realistic industrial variability.

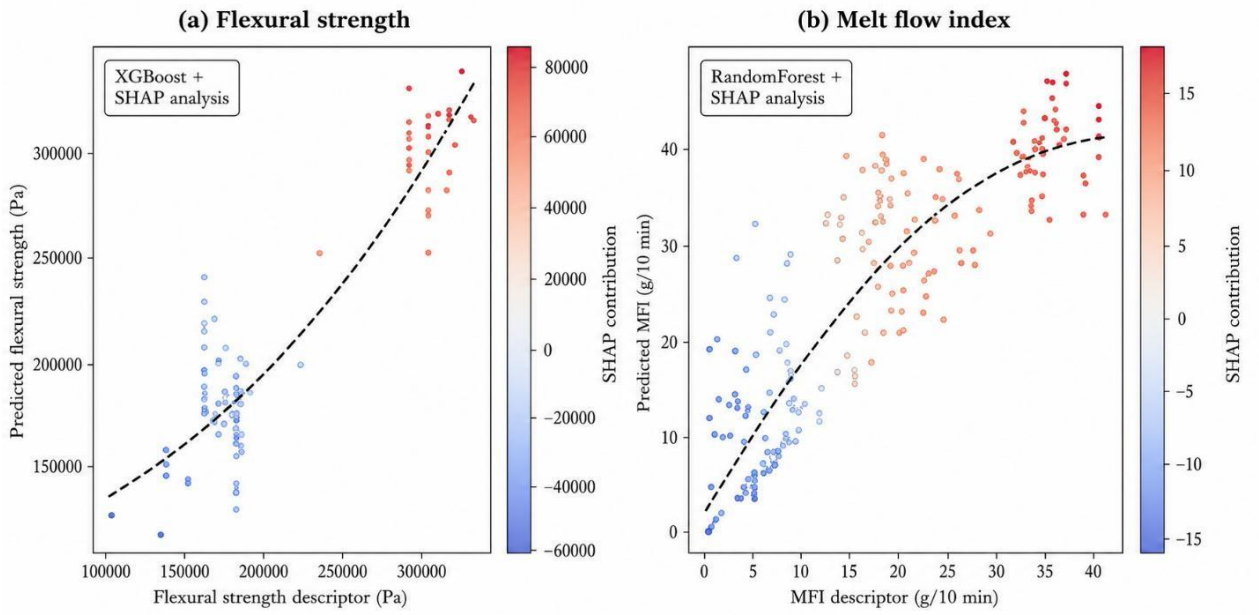

Figure 4. Descriptor–property relationships for (a) flexural strength and (b) melt flow index (MFI). Color indicates SHAP contribution, and dashed curves represent smoothed trends under industrial variability.

6. Conclusion and Outlook

This work proposes a physics-informed machine-learning framework for the prediction and interpretation of recycled polymer properties under industrially representative conditions. By restructuring experimental data into physically meaningful descriptors prior to learning, the approach enables robust prediction while preserving interpretability, even in the presence of intrinsic variability and noise characteristic of recycled materials. The framework is demonstrated on two complementary properties—melt flow index and flexural strength—spanning rheological and mechanical regimes, thereby highlighting its generality and physical consistency. Beyond predictive accuracy, the key contribution lies in the ability to interpret model behavior directly in terms of material descriptors, providing actionable insights for formulation assessment and supporting informed industrial decision-making. More broadly, this work demonstrates that integrating physics-based representations within machine-learning models enables a shift from purely predictive tools toward physically grounded and interpretable decision-support systems for recycled material design. Future work will extend this framework through deeper quantitative analysis, descriptor sensitivity studies, and broader validation across additional recycled material systems, with the objective of translating these insights into robust, deployable tools for sustainable polymer manufacturing.

7. 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, pp. 1–5, Jul. 2017, doi: 10.1126/sciadv.1700782.

[2] J. Hopewell, R. Dvorak, and E. Kosior, “Plastics recycling: Challenges and opportunities,” Philosophical Transactions of the Royal Society B, vol. 364, no. 1526, pp. 2115–2126, Jul. 2009, doi: 10.1098/rstb.2008.0311.

[3] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, “Machine learning for molecular and materials science,” Nature, vol. 559, no. 7715, pp. 547–555, Jul. 2018, doi: 10.1038/s41586-018-0337-2.

[4] S. Torquato, Random Heterogeneous Materials: Microstructure and Macroscopic Properties. New York, NY, USA: Springer, 2002.

[5] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in Proc. NeurIPS, pp. 4765–4774, 2017.

[6] K. Ragaert, L. Delva, and K. Van Geem, “Mechanical and chemical recycling of solid plastic waste,” Waste Management, vol. 69, pp. 24–58, Nov. 2017, doi: 10.1016/j.wasman.2017.07.044.

[7] F. Vilaplana and S. Karlsson, “Quality concepts for the improved use of recycled polymeric materials,” Macromolecular Materials and Engineering, vol. 293, no. 4, pp. 274–297, Apr. 2008, doi: 10.1002/mame.200700393.

[8] M. Rubinstein and R. H. Colby, Polymer Physics. Oxford, U.K.: Oxford University Press, 2003.

[9] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, 2nd ed. New York, NY, USA: Springer, 2009.

[10] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, “Physics-informed machine learning,” Nature Reviews Physics, vol. 3, pp. 422–440, Jun. 2021.