Proc. of the European Safety and Reliability Conference (ESREL 2026)
Edited by José C. Matos, Paulo B. Lourenço, Michael Beer, Daniel Oliveira,
Edoardo Patelli, Hélder S. Sousa, Rui A. Silva and Monica Santamaria
 Published by Research Publishing, Singapore.
doi: 10.3850/ESREL2026061419_esrel26-p26196-cd

AI-Driven Mutagenic Screening Tool of Plastic Monomers for Instant Safe and Sustainable by Design Assessment

António Nogueira, Juliana Rodrigues, José Ferraz-Caetano, Germán Ferreira

HOLOSS – Holistic And Ontological Solutions For Sustainability – Avenida Afonso III, S/N – Edifício do Cais da Antiga Estação da CP, da União de Freguesias de Monção e Troviscoso, 4950-431, Monção, Viana de Castelo, Portugal. E-mail: holoss@holoss.com

The use of plastics has increased over the past decades due to their versatility and widespread use in packaging, construction, and agriculture. This ubiquity results in environmental and human health challenges because of the release of plastic monomers and additives. To tackle these problems, it is important to develop novel biodegradable plastics, aligned with frameworks like the Safe and Sustainable by Design (SSbD). The SOUL project is at the forefront of these innovations, as it aims to provide bio-based solutions to replace plastics in soil applications. As such, it represents an excellent testing ground for new tools to support SSbD assessment. Aligned with SSbD, the first step should be a hazard screening, with several cut-off criteria - Mutagenicity Category 1A or 1B is amongst them. However, the process of compiling and checking mutagenicity evidence is time consuming, and data for novel chemicals is often not sufficient, hindering timely SSbD decisions. In this communication, it is presented an AI framework dedicated to evaluating the mutagenic hazard potential of plastic monomers. Using an empirically derived set of mutagenicity data, a machine learning predictive model was built that correctly classifies 92% of plastic monomers as either Mutagenic Category 1A or 1B (H340) or Mutagenic in some capacity. The developed model enables rapid prediction of mutagenicity for any plastic monomer using only its SMILES representation, requiring no experimental input. By using open-source chemical descriptors it provides a reliable go/no-go classification for mutagenic potential for plastic development within the SSbD framework. This is a user-friendly tool for carrying out early-stage risk assessment that enables stakeholders to discover hazardous compounds before product lab testing. Specifically built for plastic applications, this model will generate new insights, allowing for future evaluations of alternative materials, supporting the transition to sustainable solutions.

Keywords: risk assessment, mutagenic screening, machine learning, Safe and Sustainable by Design

1. Introduction

The global widespread and use of plastics began in the 1950s, with production increasing from near-negligible levels to millions of tons in the current times (Pilapitiya and Ratnayake 2024). They are economical, readily available, and flexible enough to meet the needs of many industries including packaging, construction and agriculture (Geyer, Jambeck, and Law 2017). Plastic products include polymers in conjunction with many other materials, which collectively provide different properties, including strength, durability, and appearance. The numerous combinations of monomers, catalysts, stabilizing agents, and plasticizers (among others) used to produce plastics leave many opportunities for undesired contaminants to be introduced (Kim *et al.* 2023). Contamination might occur in in plastic manufacturing, use, recycling, and environmental degradation. These potential pathways for contamination raise many concerns regarding the long-term implications of environmental and human health risks associated with plastic product's raw materials. (Wiesinger, Wang, and Hellweg 2021).

In 2022, Europe alone produced 58.8Mt of plastics, and generated 32.2Mt of plastic waste, further contributing to this problematic. (Plastics Europe 2024). Safe and Sustainable by Design (SSbD) reframes these concerns into an innovation workflow, where hazard identification is an early compliance filter that shapes material selection and substitution (Dekkers *et al.* 2025). In this setting, mutagenicity occupies a central role because it signals DNA-reactive potential, which is tightly coupled to carcinogenic risk (Silva *et al.* 2026). As such, hazard screening frameworks often implement explicit cut-off criteria, where classification as germ cell mutagenicity Category 1A or 1B constitutes a decisive "no-go" trigger

that can halt development or prompt redesign before further investment (Vincoff *et al.* 2024, Giri *et al.* 2025). Yet, translating this principle into routine practice remains difficult. Mutagenicity assessment typically demands a weight-of-evidence synthesis across heterogeneous test conditions, strains, and metabolic activation settings, alongside careful curation of chemical data (Ye *et al.* 2025). This process is labor intensive, and it becomes particularly restrictive for novel monomers, additives, and degradation products, for which experimental evidence is often absent. The result is a data asymmetry mismatch, where relevant mutagenicity evidence is frequently incomplete, delaying SSbD decisions and increasing the risk of developing materials that can fail hazard criteria.

To address this, data-driven mutagenicity prediction has emerged as a practical complement to experimental assessment (Lithner, Larsson, and Dave 2011). Large, public Ames test datasets now enable machine learning (ML) models that can rapidly screen chemicals and prioritize candidates for confirmatory testing (Borah and Nagamani 2024). Two methodological directions are especially relevant to SSbD use-cases: (i) global classification models optimized for broad coverage and high performance using curated descriptor sets and modern ML supervised algorithms, and (ii) similarity frameworks that embed an applicability-domain logic by cross-referencing structurally related compounds. Both these approaches suggest a pathway to accelerate early hazard screening: supporting swift go/no-go decisions for plastics-related chemicals while preserving transparency where regulatory relevance is paramount (Vogel *et al.* 2024, Shinada *et al.* 2022).

Based on this suggested early risk assessment pathway, we present an AI framework for evaluating the mutagenic hazard potential of plastic monomers. Using an empirically derived mutagenicity dataset, we developed a ML predictive model that correctly classifies with 92% accuracy if a plastic monomer is not either Mutagenic (probability of a potential Category 1A or 1B, H340 genetic event) or mutagenic in some capacity. The framework enables rapid screening of any plastic monomer directly from its SMILES (Simplified Molecular Input Line Entry System) representation, requiring no experimental input at the point of prediction. By relying on open-source chemical descriptors, the model delivers a practical go/no-go classification aligned with SSbD hazard cut-off logic, supporting early-stage material design. We focused on monomers specifically because they are the upstream chemical "decision points" of polymer design. Once selected, they constrain the chemical identity of the resulting material and influence the formation of residual products. Early prediction at the monomer stage prevents downstream lock-in to problematic chemistries and accelerates the prioritization of safer alternatives.

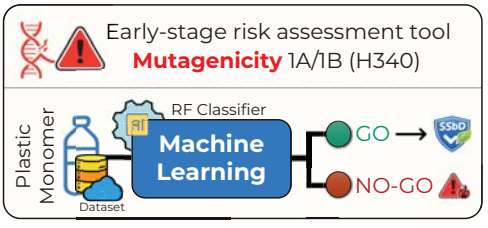

Fig. 1. Graphical summary of the ML model workflow.

We propose a plastic monomer mutagenicity screening framework by describing the data sources used to assemble the mutagenicity dataset. Next, we explain the computational process followed during model development, including how descriptors were generated, how the model was trained, and what methods were used to evaluate its predictive ability. Finally, we present and discuss the predictive performance of the resulting model, complemented by explainability assessments that clarify which molecular features drive model decisions. Together, this structure links evidence compilation, algorithmic development, and interpretable outcomes into a coherent framework suitable for early-stage SSbD hazard screening of plastic monomers.

1.1. *Testing environment: The SOUL project*

The CBE-JU SOUL project (Grant Agreement No. 101214822) is a European project aimed at developing novel biodegradable-in-soil plastics in line with the SSbD framework. These plastics will then be applied in products across various sectors, such as agriculture, sports, and landscaping.

Given the low TRL at which the project begins and the research-driven nature of the innovations, the SOUL project can serve as an environment for testing the model presented here. As SOUL's products must comply with the SSbD framework, SSbD practitioners can use the model to assess the mutagenicity of any plastic monomers developed and eliminate or replace potentially hazardous substances before product development. This approach simplifies the early evaluation of plastic monomers, reducing the need for costly in-vitro and in-vivo testing.

2. Model development for early hazard risk assessment

2.1. *Monomer database development*

The database used to develop the mutagenicity model was derived from the PlastChem inventory deposited on Zenodo (Wagner *et al.* 2024), which accompanies the report "State of the science on plastic chemicals" and provides a structured, publicly accessible compilation of plastic-associated chemicals and their hazard evidence. This source is well suited for SSbD-oriented screening because it consolidates dispersed information into a single harmonized resource and is intended to support evidence-based identification of chemicals of concern in plastics.

From the full PlastChem dataset, we extracted the subset corresponding to plastic monomers (n = 55) and retained two fields required for predictive modelling: (i) the SMILES representation (to ensure a machine-readable structural identifier) and (ii) the mutagenicity classification used for hazard screening. The mutagenicity field is encoded on an ordinal scale that reflects regulatory evidence relevance: a value of 2 indicates mutagenicity Category 1A or 1B (H340); 1 denotes that a chemical is recognized as mutagenic without specification of category; 0.5 indicates mutagenicity Category 2 (H341); 0 indicates that the chemical is not classified as carcinogenic/non-hazardous in the database context; and a blank entry denotes no data. The resulting table served as the input for open-source descriptor generation and model development.

To translate each monomer's SMILES representation into model-ready numerical features, we computed a set of 106 open-source RDKit descriptors. These descriptors provide a compact yet information-rich encoding of molecular structure, capturing complementary aspects of chemical space that are routinely informative in toxicity and property prediction tasks. In particular, they include van der Waals surface area (VSA), structural and electronic descriptors that reflect size, shape, polarity, and distribution of atomic contributions across the molecule. This descriptor set has been widely validated in prior ML studies for predicting complex chemical properties, making it a robust choice for early-stage hazard screening (Ferraz-Caetano, Teixeira and Cordeiro 2024, 2025). After data cleaning, all descriptor vectors were standardized and applied to the held-out test split.

2.2. *ML algorithm selection and optimization*

To identify a robust classification algorithm for the prediction of plastic-monomer mutagenicity, we implemented a benchmarking workflow comparing multiple supervised ML classifiers. Using near-default hyperparameters we established a baseline to select a supervised ML algorithm suited to the descriptor space. The response variable (mutagenicity) was curated to reflect regulatory phrasing and the ambiguity between low-evidence mutagenicity and non-mutagenicity. As mentioned in 2.1, four discrete mutagenicity states were defined, according to the PlastChem database (0, 0.5, 1 and 2). This discretization was adopted because the available evidence does not support a clean boundary between the 0.5 and 0.0 classes across the dataset. Consequently, two complementary learning tasks were evaluated. In the original task, the four states were retained as distinct classes. In parallel, a binary task was defined to isolate high-confidence mutagenicity signals, where labels above 0.9 (i.e., 1.0 and 2.0) were mapped to the positive class and all remaining entries (0.0 and 0.5) to the negative class. This binary operationalization reflects a pragmatic screening scenario in which strong mutagenicity evidence must be detected even when categories are not separable. From this point, we have restricted the dataset to entries with non-missing labels for the Mutagenic endpoint.

The benchmark compared four supervised ML classifiers and an additional boosting model. Random Forest (RF) and Gradient Boosting (GB) were included as non-linear tree-ensemble baselines that capture higher-order interactions between descriptors. KNN (k-Nearest Neighbors) was included as a locality-based method sensitive

to the descriptor geometry, as Logistic Regression (LR) was included as a linear baseline to quantify the extent to which the descriptor space supports linearly separable decision boundaries. XGBoost was also included as a strong regularized gradient-boosting reference commonly used for structured descriptor data. Model comparison was performed using repeated stratified holdout evaluation. For each run, the dataset was split into an 80:20 train:test split. This procedure was repeated for 50 runs with deterministic seeds to obtain distributions of performance estimates. For each split, each classifier was trained on the standardized training descriptors and evaluated on the standardized test descriptors. In addition to global test-set accuracy, the pipeline computes a domain-specific subset accuracy focused on compounds labelled as monomers. This is done to separate the other plastic components present in the PlastChem database which are not monomers.

Figure 2 presents the average results for mutagenicity prediction for the five tested ML algorithms. After the benchmarking stage, the RF classifier was selected as the best-performing algorithm across repeated stratified splits, as it had the best and second-best accuracy results for the complete and monomer-only datasets.

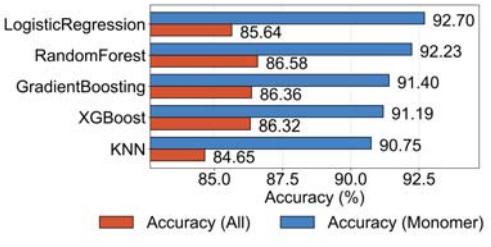

Fig. 2. Graphic summary of the evaluation accuracy for each of the tested algorithms, for the complete PlastChem dataset and for monomers only.

To improve generalization, we implemented a RF hyperparameter optimization routine using randomized search with stratified cross-validation. Two alternative learning formulations were supported: in the "original", the target labels are treated as a multi-class endpoint via label encoding. In the "binary", the model is trained to discriminate high-confidence mutagenicity (labels > 0.9, corresponding to the 1.0 and 2.0 classes) from the remaining cases (0.0 and 0.5). This binary mapping is consistent with the earlier operational decision to avoid forcing an unstable separation between low-evidence mutagenicity and non-mutagenicity. Hyperparameters are optimized exclusively on the training partition to prevent information leakage into the held-out test set. The cross-validation scheme is set to five folds, balancing variance reduction against computational cost. Model ranking during the search is based on cross-validated accuracy, allowing direct comparability between baseline and tuned performance.

2.3. *Data classification strategy*

Because the dataset includes two low-evidence states that are difficult to discriminate, mutagenic category 2 and non-mutagenic, we implemented a monomer separability analysis. The goal was to verify, under the same descriptor representation, whether the 0.5 and 0.0 groups exhibit substantial overlap that would justify treating them as partially confounded classes. The dataset is further restricted to the two classes of interest, retaining target values in {0.0, 0.5}. The resulting binary target is defined as 1 for the H341 class (0.5) and 0 for the non-mutagenic class (0.0). To quantify discriminability, two models are used: a Logistic Regression baseline and a RF model configured with parameters aligned to the main RF workflow. Both models are embedded in an identical preprocessing pipeline consisting of median imputation followed by feature standardization. For each model, cross-validated discrimination is quantified by the mean and standard deviation of ROC-AUC and average precision (AUPRC). ROC-AUC captures rank-based separability between the two classes across all decision thresholds, whereas AUPRC is more sensitive to performance imbalance, a common regime in regulatory toxicology. If the 0.5 and 0.0 groups were genuinely separable under the descriptor representation, both metrics would be expected to remain consistently above chance levels across repeats. Conversely, AUC values clustering near 0.5 and weak AUPRC values indicate that the descriptor distributions of the two groups overlap. Finally, the script includes a permutation test on mean cross-validated ROC-AUC. Class labels are permuted multiple times, and the RF ROC-AUC is recomputed under the same repeated CV protocol. The mean ROC-AUC is compared to the null distribution under permutation, yielding a p-value as the fraction of

permuted scores. If the p-value is non-negligible and the observed AUC is close to the permutation mean, the result supports that the RF is not learning a reproducible separation.

In the monomer subset (n = 55; 32 in the 0.5/H341 class and 23 in the 0.0/non-mutagenic class), we evaluated whether the two low-evidence categories are separable under the same descriptor representation used in the main workflow. Across repeated 5-fold cross-validation, both a linear baseline (logistic regression) and a non-linear model (random forest) yielded discrimination close to, or below, chance. Logistic regression achieved a mean ROC-AUC of 0.385 (SD 0.132) with an AUPRC of 0.578 (SD 0.100), essentially matching the class-prevalence baseline (~0.58) and indicating no meaningful ranking advantage. RF performed with a ROC-AUC of 0.468 and AUPRC of 0.618, as the permutation test further supported the absence of signal (permuted mean AUC 0.537), consistent with overlap between the 0.5 and 0.0 groups. These results indicate that monomers with the 0.5 (mutagenic category 2) and 0.0 (non-mutagenic) labels are not distinguishable, justifying their treatment as partially confounded.

2.4. *Model tuning and optimization*

The final modelling script was designed to operationalize mutagenicity prediction. The script evaluates Mutagenicity and repeats the full train–test evaluation procedure across 50 stratified splits, using a test fraction of 0.2. In binary determination mode, the model is reframed as a high-confidence mutagenicity detector by thresholding the numeric target at 0.9, assigning the positive class to entries labelled 1.0 or 2.0 and the negative class to entries labelled 0.0 or 0.5. This binary construction is consistent with the earlier evidence that the 0.5 and 0.0 groups overlap substantially in monomer-only feature space. For each run, descriptors are standardized, as the pipeline applies embedded feature selection. The selected feature subset is tracked explicitly to ensure downstream feature-importance. The predictive RF classifier was configured with the tuned parameterization, as in each run the model is trained on the training subset and then used to predict labels for the held-out test subset.

In binary mode, the script further expands evaluation, reported as test-set percentages of true negatives, false positives, false negatives, and true positives. This is particularly relevant for mutagenicity screening, where false negatives may be more critical than false positives depending on the decision context. The script provides a functional stratification of prediction behaviour from its metadata, as each monomer test entry may be associated with one or more functional group or mechanism tags. The code parses these tags, and in binary mode it accumulates per-function counts of TN/FP/FN/TP outcomes. The resulting stratification supports interpretation of whether the model behaves consistently across chemically subdomains or its error modes concentrate in classes.

To support model explainability, the script records two complementary feature attribution outputs during binary evaluation. First, it stores the RF impurity-based feature importances for the selected feature subset at each run, yielding a distribution of importance values across splits. Second, it computes SHAP (SHapley Additive exPlanations) values using the held-out test set and extracts SHAP contributions corresponding to the positive class. SHAP values are aggregated for the monomer-only subset of the test set by computing the mean absolute SHAP magnitude per feature, producing an interpretable ranking of the descriptors most responsible for positive-class assignments.

3. Model results for risk assessment

3.1. *Accuracy assessment*

The classifier model achieved high accuracy for monomer classification (≈92.0%) and exhibited strong specificity, reflected by a low false-positive rate (False Positive ≈0.22%) and a high true-negative proportion (True Negative ≈91.8%). Figure 3 presents the confusion matrix for the monomer classification. This further supports interpretation of the model as a conservative risk-screening tool: negative predictions can be used to deprioritize compounds unlikely to belong to the confirmed mutagenic class, whereas positive predictions should be treated as flags requiring confirmatory testing rather than as definitive calls.

In a go/no-go framing, negatives support a provisional “go” (advance/ deprioritize for testing), while positives trigger a “no-go” pending confirmatory assays.

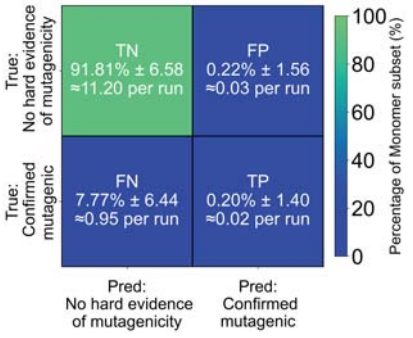

Fig. 3. Confusion matrix for the RF monomer classification (TN = True Negative; FP = False Positive; FN = False Negative; TP = True Positive). The colormap displays the percentage of monomers.

As Figure 3 shows, sensitivity to the mutagenic class was limited, consistent with low prevalence of Mutagenic > 0.9 instances available for model learning. Thus, the model is best positioned for early risk assessment aimed at ruling out high mutagenicity, rather than as a stand-alone detector of confirmed mutagenic compounds.

A key limitation of the present binary formulation is the low number (and low prevalence) of confirmed mutagenic instances (Mutagenic > 0.9). This class imbalance reduces the effective sample size available for learning the positive class decision boundary and increases the variance of positive-class performance estimates across splits. Consequently, the classifier exhibits strong specificity (few false positives) but limited sensitivity for the mutagenic class. In this context, the model is best interpreted as an early-stage risk-screening tool for deprioritizing compounds unlikely to be strongly mutagenic, while positive predictions should be treated as flags requiring confirmatory testing.

Functional stratification by metadata tags shows that, among the categories with substantial support (those appearing more than 600 times), predictive accuracy remains consistently high and varies within a relatively narrow band. As shown in Figure 4, in this high-frequency subset, the strongest performance is observed for Monomer (mean accuracy 88.6%), followed by Antioxidant (85.9%), Biocide (85.9%), Colorant (85.4%), and Odor agent (85.1%).

These results indicate that, when sufficient tagged evidence is available, the model generalizes most reliably for these functionally defined regimes, supporting the use of metadata-based stratification to identify well-characterized domains of robust predictive behavior.

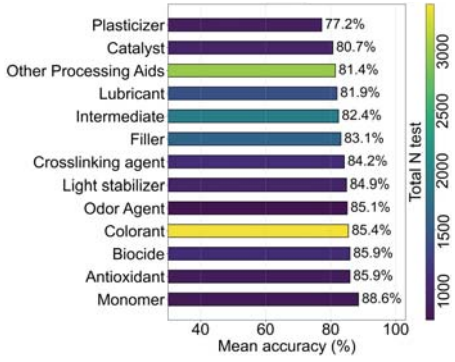

Fig. 4. Monomer stratification by model prediction accuracy for N > 600 functional tags in the dataset.

3.2. *Descriptor Explainability*

To obtain a global view of how different descriptor families contribute to prediction, we aggregated the results of two complementary explainability approaches: feature importance and SHAP across all three descriptor types. This dual analysis is useful because it contrasts a ranking-based notion of "importance" of how features drive the model output over the dataset. Figure 5 presents the overall average results for both metrics for each descriptor type. Figure 5 shows that feature importance distributed across

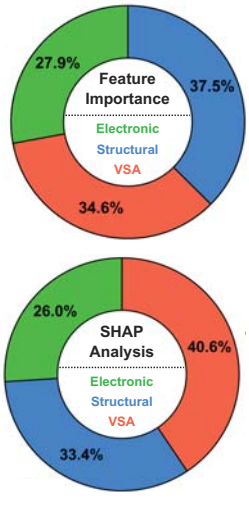

Fig. 5. Average results for feature importance and SHAP analysis after model prediction, grouped by descriptor types.

descriptor families rather than dominated by a single type. Structural descriptors contribute the largest share (37.5%), VSA descriptors are nearly as influential (34.6%), and electronic descriptors remain substantial (27.9%). This balance suggests the model is using multiple, complementary chemical "views" (connectivity plus surface-property proxies, with added electronic effects) rather than relying on one narrow signal. For explainability, this pattern implies that interpretations should be framed at the descriptor-family level rather than attributing predictions to a single decisive feature. The individual importances do not indicate a single dominant feature. Even the high ranked descriptors (as shown in Appendix A) contribute only modestly, and the remaining top variables cluster around similar magnitudes. The absence of a large gap between the top features supports explainability as a multi-factor pattern.

SHAP assigns each feature a contribution to each individual prediction, grounded in Shapley values from cooperative game theory. When summarized, it reflects the average magnitude of a feature's effect on the model output. In the present results, SHAP indicates that VSA descriptors contribute the largest share of predictive impact (≈40.6%), followed by Structural descriptors (≈33.4%), while Electronic descriptors contribute the smallest share (≈26.0%). This pattern aligns closely with the overall feature-importance analysis. Taken together, both analyses converge on the same conclusion: VSA and Structural information primarily govern the model's predictive behavior, with Electronic descriptors providing a secondary but non-negligible contribution.

This explainability feature is a key advantage against recent ML mutagenicity alternatives. Close alternatives is Ames test ML prediction modelling, which rely on strategies that are less tied to explicit chemical features: either distance-based applicability domain criterion (Li, Liu, Thakkar, Roberts, Tong 2023) or using molecular images processed by a convolutional neural network (Tran, Tayara, and Chong 2024). Although these approaches can achieve strong predictive performance, they provide less insight into which chemically meaningful properties drive the prediction. This gives our model greater chemical interpretability and makes the predictions easier to relate to known determinants of mutagenic behavior. However, direct comparison remains limited, as these models were evaluated on often non-plastic test sets with different performance metrics.

4. Final Remarks

This article demonstrates that an SSbD-aligned monomer screening AI (ML) model can be constructed from harmonized mutagenicity labels and standard RDKit descriptors, yielding a conservative classifier that strongly rules out high-confidence mutagenicity. With an overall monomer-level accuracy of 92%, the RF classification model can be operationalized as a go/no-go screening tool: supporting "go" decisions to deprioritize candidates with low predicted mutagenicity risk, while assigning "no-go" flags to positives that warrant confirmatory testing. The key interpretability result is that prediction is not driven by a single chemical signal: both feature importance and SHAP converge on a distributed reliance across descriptor families, with VSA and structural information dominating and electronic descriptors providing a consistent contribution.

For risk assessment, these characteristics position the model as an early-stage screening layer by flagging candidates that fall into the mutagenic class. It is most useful upstream of regulatory or experimental workflows where large chemical inventories must be narrowed under limited testing capacity (as for polymer design or portfolio screening). At the same time, limited sensitivity reflects the scarcity and imbalance of confirmed mutagenic examples. The demonstrated non-separability between the 0.5 and 0.0 states also indicates that finer-grained categorization is data limited. Potential improvements therefore include enlarging the positive class (high-quality mutagenicity datasets), calibrated to false-negative risk with expanded descriptor coverage. In this way, the model can evolve into a decision-support tool.

When applied to the SOUL project the model can accelerate early-stage safety testing of the developed plastics, improving the safety aspects of the final products. Using SOUL as a testing environment can also provide the model with experimental data, further validating the AI predictions and guiding the future of monomer safety assessments.

Acknowledgement

Dedication and funding information may also be included here.

Author Contributions

A. Nogueira, J. Rodrigues: Writing – original draft, review & editing. *J. Ferraz-Caetano*: Writing – original draft, review & editing; Visualization; Software; Methodology; Data curation. *Germán Ferreira*: Writing – review & editing; Methodology.

Appendix A. Online Code and Data Repository

The data, code and full statistical results can be found online at https://github.com/jfcaetano/PlastMuta.

References

Borah, G. S. and S. Nagamani (2024). Development of a robust machine learning model for Ames test outcome prediction. Chemical Physics Letters 856, 141663.

Dekkers, S., V. Adam, V. Di Battista, A. Haase, J. Prinz, G. Nagel, W. Fransman, *et al.* (2025). Safe-and-sustainable-by-design approach and decision support system for advanced materials. Advanced Sustainable Systems 9(10), e00208.

Ferraz-Caetano, J., F. Teixeira, and M. N. D. S. Cordeiro (2025). Optimizing vanadium-catalyzed epoxidation reactions: Machine-learning-driven yield predictions and data augmentation. Journal of Chemical Information and Modeling 65(13), 6757–6771.

Ferraz-Caetano, J., F. Teixeira, and M. N. D. S. Cordeiro (2024). Data-driven, explainable machine learning model for predicting volatile organic compounds' standard vaporization enthalpy. Chemosphere 359, 142257.

Geyer, R., J. R. Jambeck, and K. L. Law (2017). Production, use, and fate of all plastics ever made. Science Advances 3(7), e1700782.

Giri, S., G. Lamichhane, J. Pandey, and D. Khadka (2025). Micro(nano)plastics in human carcinogenesis: Emerging evidence and mechanistic insights. Microplastics 4(4), 78.

Kim, M. S., H. Chang, L. Zheng, Q. Yan, B. F. Pfleger, J. Klier, K. Nelson, E. L.-W. Majumder, and G. W. Huber (2023). A review of biodegradable plastics: Chemistry, applications, properties, and future research needs. Chemical Reviews 123(16).

Li, T., Z. Liu, S. Thakkar, R. Roberts, and W. Tong (2023). DeepAmes: A deep learning-powered Ames test predictive model with potential for regulatory application. Regulatory Toxicology and Pharmacology 144, 105486.

Lithner, D., Å. Larsson, and G. Dave (2011). Environmental and health hazard ranking and assessment of plastic polymers based on chemical composition. Science of the Total Environment 409(18), 3309–3324.

Nayanathara Thathsarani Pilapitiya, P. G. C., & Ratnayake, A. S. (2024). The world of plastic waste: A review. Cleaner Materials, 11, 100220.

Plastics, Europe (2024) The Circular Economy for Plastics – A European Analysis 2024. https://plasticseurope.org/knowledge-hub/the-circular-economy-for-plastics-a-european-analysis-2024/ (accessed February 25, 2026)

Shinada, N. K., N. Koyama, M. Ikemori, T. Nishioka, S. Hitaoka, A. Hakura, S. Asakura, Y. Matsuoka, and S. K. Palaniappan (2022). Optimizing machine-learning models for mutagenicity prediction through better feature selection. Mutagenesis 37(3–4), 191–202.

Silva, G. C. A., M. N. Fioretto, M. C. Lima, W. R. Scarano, and F. K. Delella (2026). Micro and nanoplastics in human carcinogenesis: Insights from in vitro studies. Toxicology 519, 154331.

Trairatphisan, P., L. Dorsheimer, P. Monecke, J. Wenzel, R. James, A. Czich, Y. Dietz-Baum, and F. Schmidt (2025). Machine learning enhances genotoxicity assessment using MultiFlow® DNA damage assay. Environmental and Molecular Mutagenesis 66(1–2), 45–57.

Tran, T. T., H. Tayara, and K. T. Chong (2024). AMPred-CNN: Ames mutagenicity prediction model based on convolutional neural networks. Computers in Biology and Medicine 176, 108560.

Vincoff, S., B. Schleupner, J. Santos, M. Morrison, N. Zhang, M. M. Dunphy-Daly, W. C. Eward, A. J. Armstrong, Z. Diana, and J. A. Somarelli (2024). The known and unknown: Investigating the carcinogenic potential of plastic additives. Environmental Science & Technology 58(24).

Vogel, A., J. Tentschert, R. Pieters, F. Bennet, H. Dirven, A. van den Berg, E. Lenssen, M. Rietdijk, D. Broßell, and A. Haase (2024). Towards a risk assessment framework for micro- and nanoplastic particles for human health. Particle and Fibre Toxicology 21, 48.

Wagner, M., L. Monclús, H. P. H. Arp, K. J. Groh, M. E. Løseth, J. Muncke, Z. Wang, R. Wolf, and L. Zimmermann (2024). State of the science on plastic chemicals - Identifying and addressing chemicals and polymers of concern. Zenodo.

Wiesinger, H., Z. Wang, and S. Hellweg (2021). Deep dive into plastic monomers, additives, and processing aids. Environmental Science & Technology 55(13), 9339–9351.

Ye, S., J. Xiang, S. Zhou, Q. Ge, A. Anwaier, K. Chang, G. Wei, et al. (2025). Polystyrene microplastics exposure aggravates clear cell renal cell carcinoma progression via the NF-κB and TGF-β signaling pathways. Advanced Science (published online November 27).