Quantifying Algorithmic Decision-Making in Engineering: A High-Fidelity Explainable AI (XAI) Framework for Non-Linear Regression

Wing Lam Jasmine Tsang*
*First and Corresponding author, Independent Researcher,
wl.tsang13@alumni.imperial.ac.uk

Abstract

Applying machine learning to safety-critical materials science is often limited by the blackbox nature of AI models. This study presents an explainable AI (XAI) framework to balance predictive accuracy with physical interpretability. There are four distinct architectures (Linear Regression, Support Vector Regression (SVR), Gradient Boosting, and Random Forest) in this study, using a Test Split R^2 and 10-fold Cross-Validation to ensure results are generalisable. SHapley Additive exPlanations (SHAP) method is used to study the effect of different factors that can affect materials performance. This approach follows Green AI concept, without the requirements of resource-intensive laboratory sessions, it provides the same prediction values.

This study successfully identified the significant effect of fracture energy under low temperature testing, where it showed a sharp sensitivity threshold at -40 °C. The use of explainable AI (XAI) helps to provide more insights and practical predictions for further engineered designs and enhancements. It also reduced the use of non-recyclable materials required in laboratory. This study showed that by using explainable AI models, even with limited experimental results can provide additional insights and practical predictions for further engineered designs and enhancements.

1.Introduction

The development of artificial intelligence (AI) and Machine Learning (ML) have greatly enhanced engineering research work. While the use of Machine Learning (ML) can be very useful in predicting mechanical properties, the model is limited by its black box nature[1]. Unlike traditional black-box model, explainable AI (XAI) approach[2, 3] works with real world experimental data and focus on how accurate the predicted results are when compared to real life results.

By interpreting though SHAP, we can use the model for checking and improving engineering designs, especially regarding the specific combination of mechanical properties and conditions requirements[4, 5].

The experimental data in the study are from PhD thesis of W.L. Tsang, Imperial College London [6, 7]. This PhD work characterised the mechanical and fracture performance of an anhydride-cured thermosetting epoxy polymer modified with various weight percentages (wt%) of 20-nm silica nanoparticles (ranging from 0.5 to 25.4 wt%), pre-formed core-shell rubber (CSR) particles of between 106 μm and 38 μm (ranging from 0.5 to 10 wt%), and hybrid formulations containing equal weight percentages of both particles. The dataset includes laboratory measurements of mechanical properties such as the tensile strength, fracture toughness and fracture energy. The testing was performed at ambient temperature as well as low temperatures of -40°C and -80°C. When the testing was performed in low temperature conditions, the different toughening mechanisms in brittle fracture conditions were revealed. There were also different speed of mechanical testing studies performed. All these extreme brittle fracture testing conditions are very useful in predicting situations in car or plane crush and hence helped in optimising designs for safety measurements.

SHAP (SHapley Additive exPlanations) is a common technique in explainable AI (XAI), other models such as LIME (Local Interpretable Model-Agnostic Explanations)[8] have also showed improve in interpretability of AI model.[9, 10] There are many challenges in applying explainable AI (XAI) models, for example there can be absence of standardisation, limitations with user understanding and domain specific challenges in the study too[11]. Therefore, working with lab based experimental results can help to ensure that performances are realistic to real life environments.

2. Methodology: The XAI Framework

2.1 Algorithmic Architecture and Preprocessing

Preliminary sensitivity analysis revealed that Core-Shell Rubber (CSR) modified epoxy samples introduced significant variance, as their micro-scale nature fundamentally differs from the nano-scale silica particles, leading to a marked decrease in R^2 performance. Although the effect can be accounted with calculations, there are bigger micro sizes of Core-Shell Rubber (CSR) where their toughening effect varies and hence would produce inaccurate results. Therefore, there is a preprocessing of data that involved removal of some Core-Shell Rubber (CSR) modified epoxy data points, the dataset consists of 20 samples. This is for removing outliners. The input features are following the physical parameters (Temperature and Particle Content) prior to Random Forest training.

This approach follows Green AI concept, without the requirements of resource-intensive laboratory sessions, it provides the same prediction values.

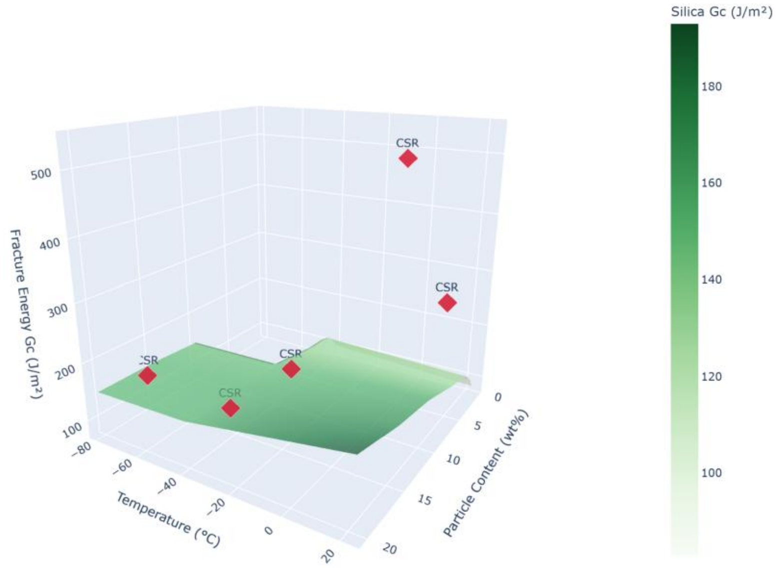

Figure 1Comparative analysis showing the divergence in fracture energy patterns between silica nanoparticles and CSR microparticles, justifying the filtering of CSR as an outlier for this specific AI logic.

2.2. Algorithmic Benchmarking and Model Selection

In order to ensure the selection of architecture is optimal for the explainable AI (XAI) model, a benchmarking study was conducted. The dataset was evaluated into four algorithmic models:

1. **Linear Regression:** It presents the linear relationship between the different factors with the fracture energy (G_c).

2. **Support Vector Regression (SVR):** It is an extension of Support Vector Machines (SVM) that is designed to solve regression problems. It works by introducing a defined region of tolerance around the function to optimise. It is for finding the boundary that best approximates the continuous-valued function while minimising the prediction error. This is also the direct difference between the predicted value and the true target value. [12]

3. **Optimised Random Forest:** Random Forest is a flexible learning model capable of tackling both regression and classification problems. It operates by constructing multiple "decision trees" during the training period and producing an average forecast from all the trees involved. [13]

4. **Gradient Boosting:** It is another type of ensemble supervised machine learning (ML) that is used for both classification and regression problems. Techniques such as random forests and decision trees. They are called ensemble because they combine results from numerous individual models for the final result model. [14]

The models are evaluated with both training R^2 and 10-Fold Cross-Validation R^2. It is for the identification of the architecture that fit the best for the specific conditions in the real physical situation, and also for SHAP analysis.

2.3 Pearson Analysis

The dataset focused on the relationship between testing temperature, silica particle content (wt%), and fracture energy (G_c). Unlike nano-silica, CSR micro-particles toughening mechanisms including cavitation and shear-banding at different fracture energy values, and this has been taken into account.

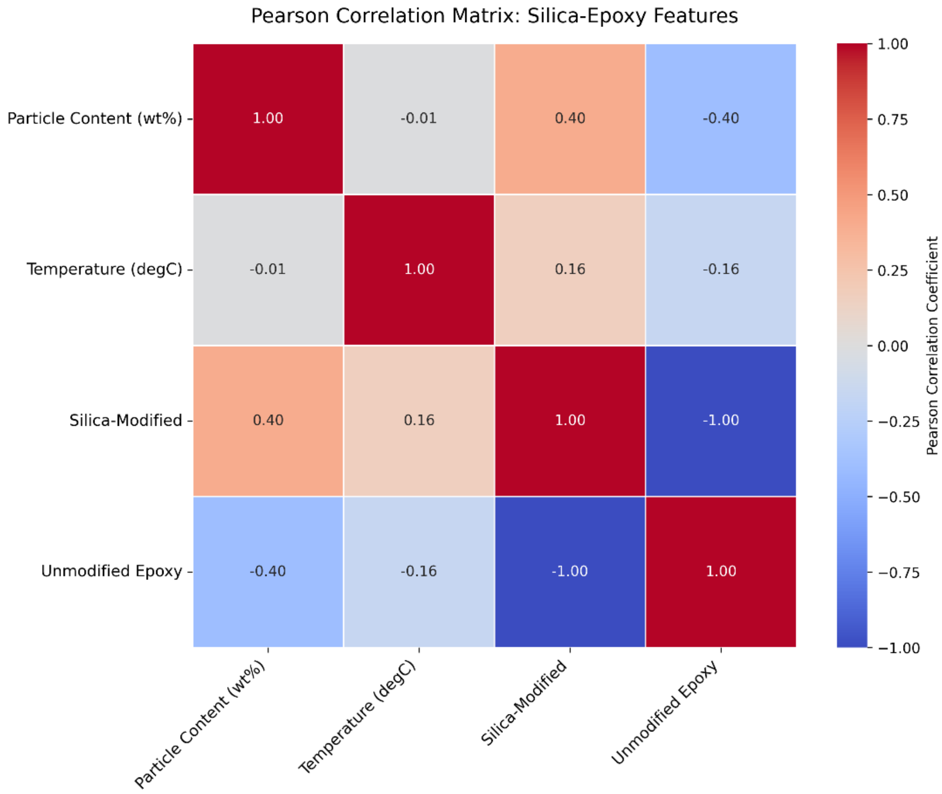

Figure 2 Pearson Correlation Matrix showing linear coefficients between environmental variables and fracture energy.

3. Results and Discussion

3.1. Algorithm comparisons results

Algorithmic comparisons of the four algorithmic models (Linear Regression, Support Vector Regression (SVR), Optimised Random Forest and Gradient Boosting), comparing performance in between experimental results and predicted results:

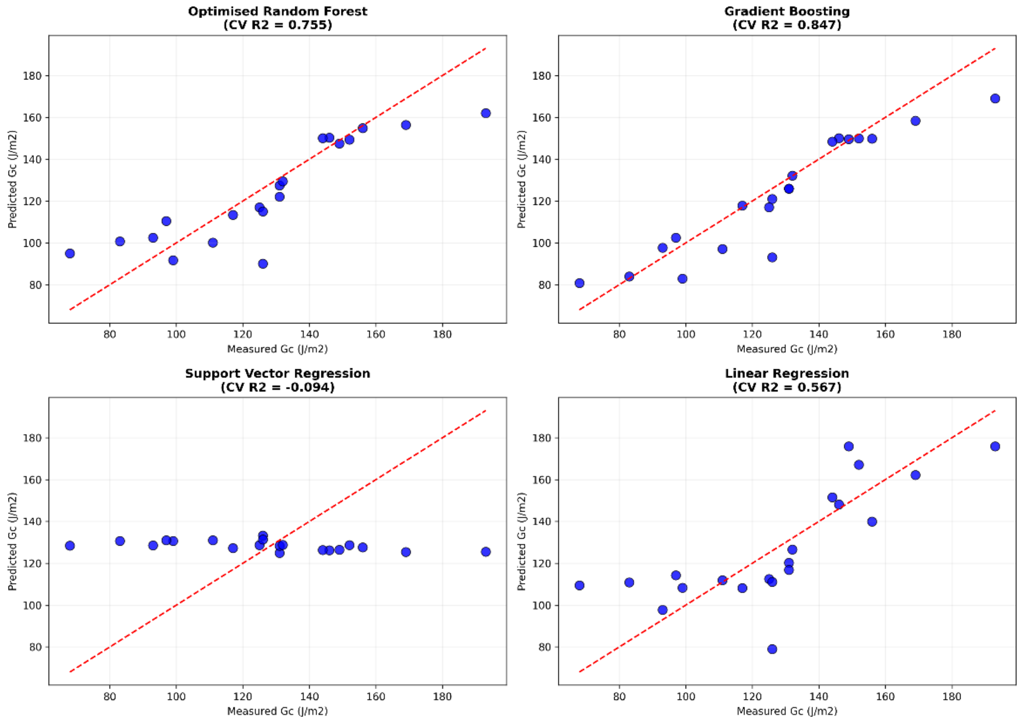

Figure 3 Comparative scatter plots demonstrating predicted versus actual Gc (J/m²) across Linear Regression, Support Vector Regression, Optimised Random Forest, and Random Forest architectures.

Gradient Boosting and Optimised Random Forest models have the highest Cross-Validation (CV R^2) when compared to Linear Regression and Support Vector Regression (SVR). When compared to the ideal y = x dashed reference line, Gradient Boosting model resulted in cross-validation coefficients of determination of 0.847 (CV R^2 = 0.847), and Optimised Random Forest models resulted in cross-validation coefficients of determination of 0.755 (CV R^2 = 0.755).

Linear Regression baseline resulted in cross-validation coefficients of determination of 0.567 (CV R^2 = 0.567), it showed that linear regression cannot capture all the features in the toughening mechanisms and related mechanical properties effects.

Support Vector Regression (SVR) model resulted in cross-validation coefficients of determination of -0.094 (CV R^2 = -0.094), it is not the correct model for this relationship, it cannot be used for explaining true measured experimental values. Therefore, it is not a suitable model for these experiments.

3.2 Algorithmic Benchmarking and the account for Overfitting

The performance metrics in Table 1 showed significant variances in how the experimental noise from real life data has an effect on the model.

Model	10-Fold CV R^2	Test Split R^2	Mean Absolute Error (MAE) (J/m^2)	Root Mean Squared Error (RMSE) (J/m^2)
Optimised Random Forest	**0.755**	**0.688**	**13.97**	**18.29**
Gradient Boosting	0.847	0.440	17.72	24.52
Support Vector Regression	-0.094	-0.271	30.67	36.94
Linear Regression	0.567	-0.220	29.42	36.19

Table 1 Table showing the 10-Fold CV R^2 and Test Split R^2 of the four models with Mean Absolute Error (MAE) (J/m^2) and Root Mean Squared Error (RMSE) (J/m^2)

Optimised Random Forest: Optimised Random Forest has the most stable values among all four models presented. Although when compared to Gradient Boosting, it has a slightly smaller 10-Fold CV R^2 value (0.755), it has a significantly higher Test Split R^2 value (0.688), with more stable Mean Absolute Error (MAE) (13.97) and Root Mean Squared Error (RMSE) (18.29). Therefore, it is the most suitable model for this study.

Gradient Boosting: Gradient Boosting has the highest 10-Fold CV R^2 value, but a low Test Split R^2 (0.440), Mean Absolute Error (MAE) (17.72) and Root Mean Squared Error (RMSE) (24.52) are not too high when compare to other models, but the low Test Split R^2 (0.440) showed overfitting, which means the model cannot be correct when it is used in a wider range of experiment data.

Support Vector Regression: Support Vector Regression showed negative results in 10-Fold CV R^2 (-0.094) and Test Split R^2 (-0.271), which means they are not showing correct relationships, and hence cannot be used.

Linear Regression: Linear Regression gave a reasonable 10-Fold CV R^2 value (0.567), but a negative Test Split R^2 (-0.220). The Mean Absolute Error (MAE) (29.42) and Root Mean Squared Error (RMSE) (36.19) are high when compare to other models, and with the low Test Split R^2 value (0.440) showed that results are not stable there. Hence, linear regression cannot be the correct relationship there. This is also showed in the comparative scatter plot (Figure 3).

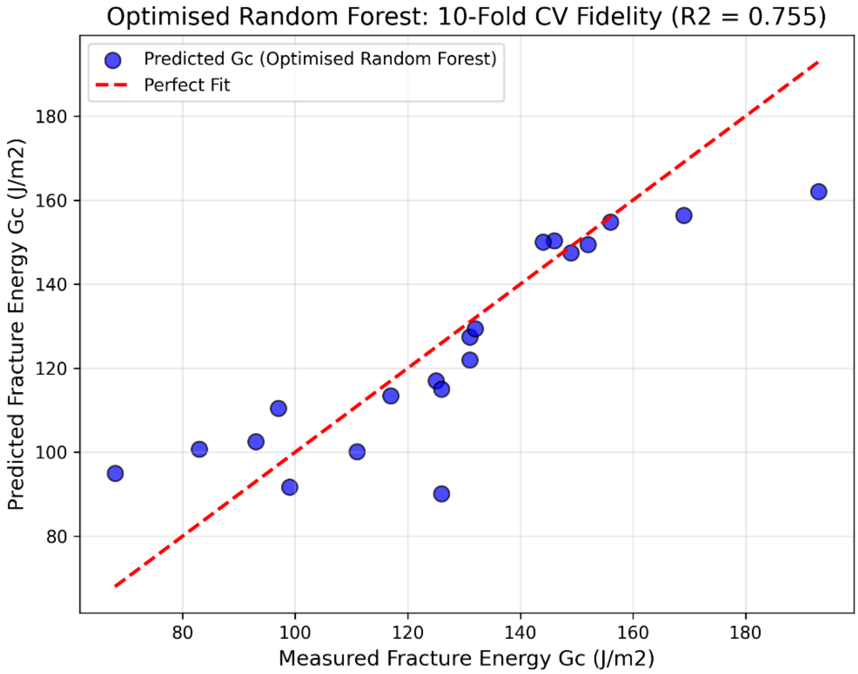

Figure 4 Optimised Random Forest (10-fold CV) for Predicted fracture energy vs measured fracture energy.

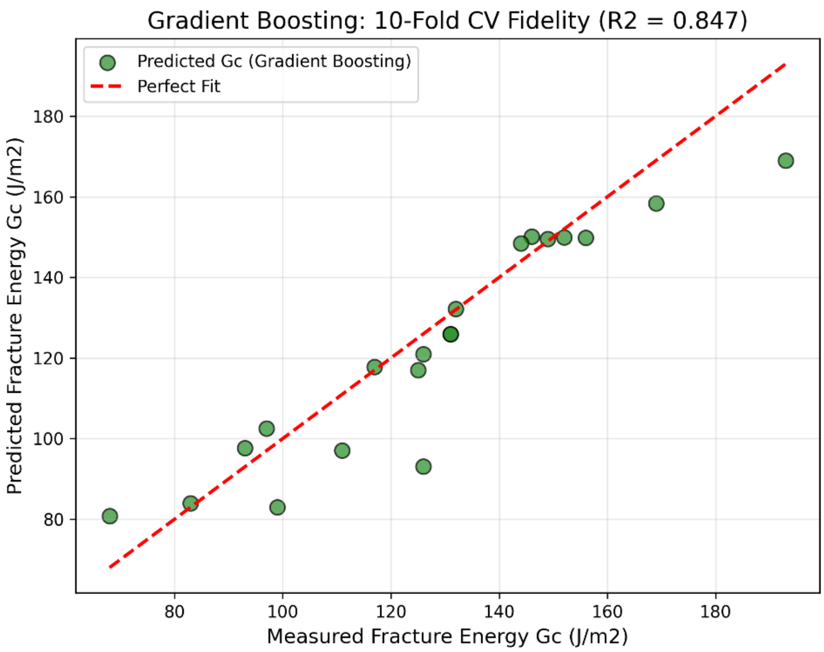

Figure 5 Gradient Boosting (10-fold CV) for Predicted fracture energy vs measured fracture energy

When compare the results from figure 4 and 5 to other studies, similar R square values are found in the study from Tonin et al[15], where they compare the R square value of Extreme Gradient Boosting (XGBoost), Random Forest (RF), Penalized Regression (PR) and Neural Network (NN)[15].

3.3. Global Feature Attribution of each factor

A Global SHAP Dashboard is used to show the effect of the factors and their impact on the model outcome. This analysis considers each material property as a mathematical factor and compares their contribution value to each other. Temperature and weight percentage of silica particles are found to be the most dominant factors to fracture energy values.

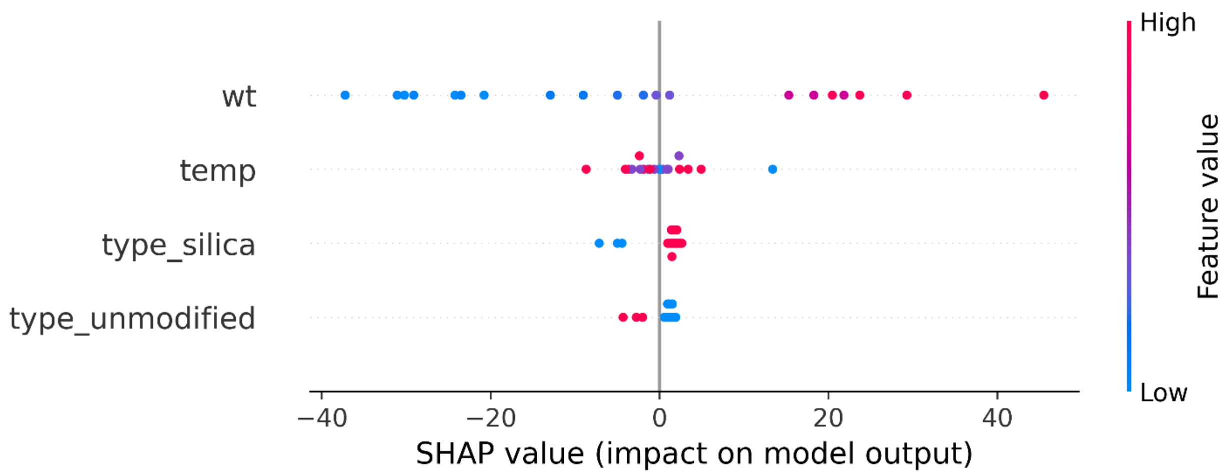

Figure 6 Global SHAP Summary Plot.

3.4. Physical Validation: Saturation and Threshold Analysis

Using SHAP Dependence plots, we can investigate specific variables that have significant effect on fracture energy (G_c) when all other factors are unchanged. The effect of different factors and how the variables can affect performance can be viewed and analysed with the specific combinations and situations.

The -40°C temperature Point: There is a sharp slope change below -40°C, corresponding to changes found in toughening mechanism.
The Toughening effect Plateau: The model identifies a saturation point around 15 wt% for silica particles, which matches experimental findings regarding silica nano particles (20nm in diameter) debonded and followed by plastic void growth.

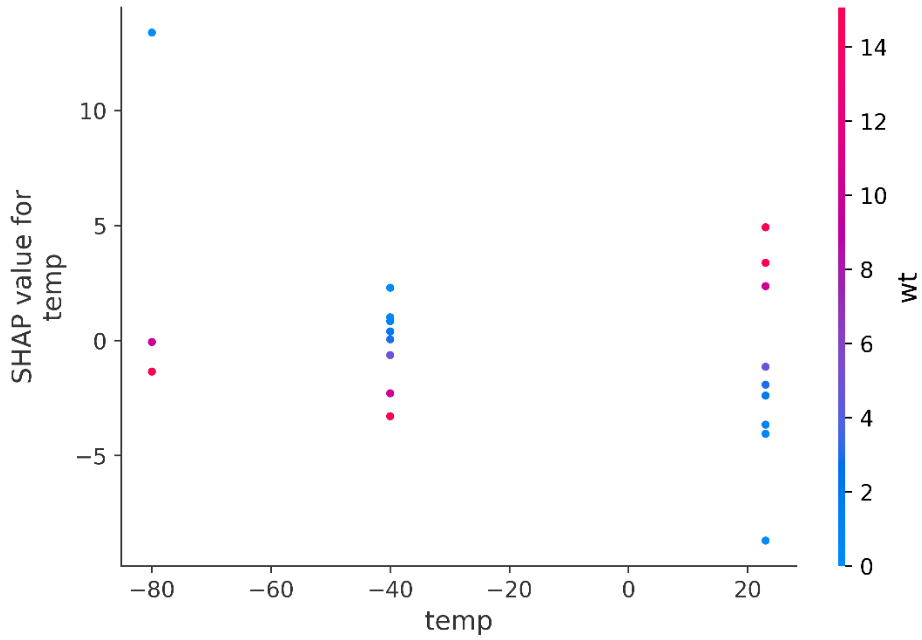

Figure 7 Dependence plot for Temperature, highlighting the critical transition at -40°C.

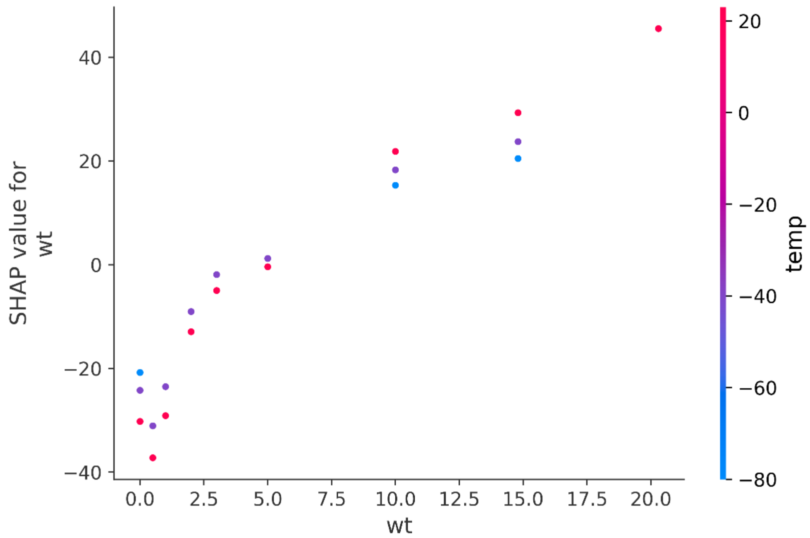

Figure 8 Dependence plot for Particle Weight % with the SHAP value

3.5. Mapping and visualising the Design Space

In order to visualise the interaction between temperature and silica on fracture energy, a 2D heatmap and an Interactive 3D Fracture Energy model are used. This heatmap provides mechanical properties for designed performance without requiring further experimental testing.

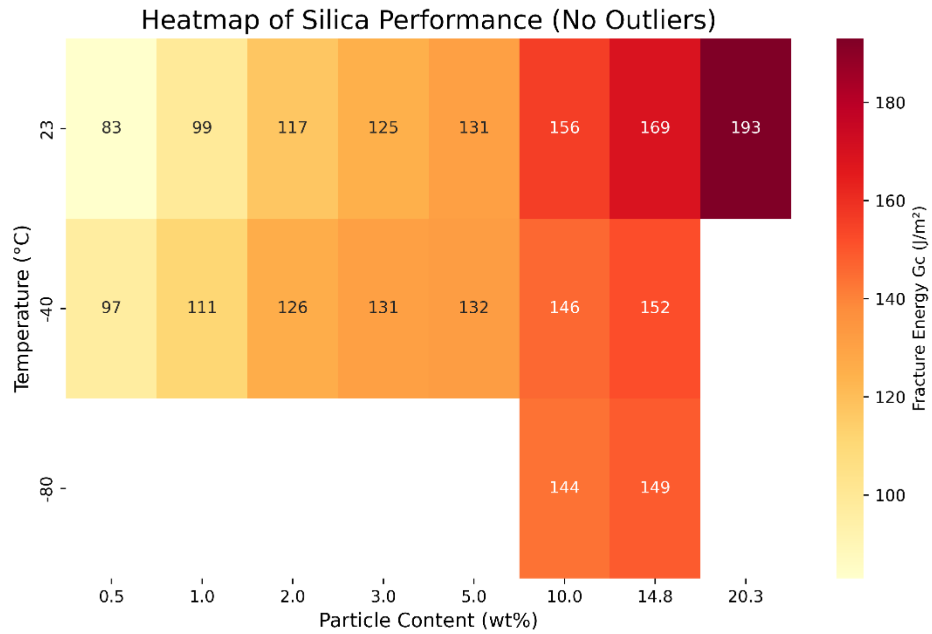

Figure 9 2D Heatmap visualising the interaction between temperature and silica on fracture energy.

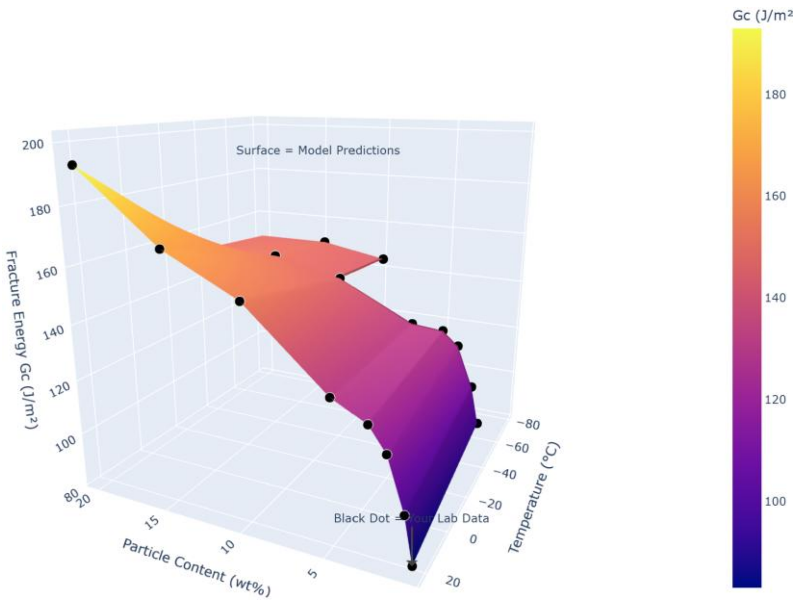

Figure 10 Interactive 3D Fracture Energy model.

4. Conclusion

This study establishes a roadmap for more environmental friendly material science AI research studies, where results are not based on the blackbox assumptions, but real-life data.

By using explainable AI models, even with limited experimental results can provide additional insights and practical predictions for further engineered designs and enhancements.

It also reduced the use of non-recyclable materials required in laboratory work.

The preliminary algorithmic benchmarking model proved that complicated models are not always optimal for all materials informatics. The model framework from this study showed that interpretability would be one of the key factors in the use of AI in related engineering discoveries.

Data Availability Statement: The Python code and dataset supporting this study are available on Zenodo at: https://doi.org/10.5281/zenodo.21733591

The author declares no competing interests.

This research received no funding from any funding agency in the public, commercial, or not-for-profit sectors.

1. Liu, B., Liu, P., Lu, W. and Olofsson, T. (2025), Explainable Artificial Intelligence (XAI) for Material Design and Engineering Applications: A Quantitative Computational Framework. International Journal of Mechanical System Dynamics, 5: 236-265. https://doi.org/10.1002/msd2.70017.
2. Dwivedi, R., et al., *Explainable AI (XAI): Core Ideas, Techniques, and Solutions.* ACM COMPUTING SURVEYS, 2023. **55**(9).
3. Bennetot, A., et al., *A Practical Tutorial on Explainable AI Techniques.* ACM COMPUTING SURVEYS, 2025. **57**(2).
4. Holzinger, A., et al., *Explainable AI Methods - A Brief Overview*, in *xxAI - Beyond Explainable AI: International Workshop, Held in Conjunction with ICML 2020, July 18, 2020, Vienna, Austria, Revised and Extended Papers*, A. Holzinger, et al., Editors. 2022, Springer International Publishing: Cham. p. 13–38.
5. Zhang, S., et al., *Fatigue life analysis of high-strength bolts based on machine learning method and SHapley Additive exPlanations (SHAP) approach.* Structures, 2023.
6. Tsang, W.L. and A.C. Taylor, *Fracture and toughening mechanisms of silica- and core–shell rubber-toughened epoxy at ambient and low temperature.* Journal of Materials Science, 2019. **54**(22): p. 13938–13958.
7. Tsang, W.L., *The use of tapered double cantilever beam (TDCB) in investigating fracture properties of particles modified epoxy.* SN Applied Sciences, 2020. **2**(4): p. 751.
8. Givisis, I., et al. *Comparing Explainable AI Models: SHAP, LIME, and Their Role in Electric Field Strength Prediction over Urban Areas*. Electronics, 2025. **14**, 4766 DOI: 10.3390/electronics14234766.
9. Gunning, D. and D.W. Aha, *DARPA's Explainable Artificial Intelligence Program.* AI Magazine, 2019. **40**(2): p. 44–58.
10. Gaspar, D., P. Silva, and C. Silva, *Explainable AI for Intrusion Detection Systems: LIME and SHAP Applicability on Multi-Layer Perceptron.* IEEE Access, 2024. **PP**: p. 1–1.
11. Kalasampath, K., et al., *A Literature Review on Applications of Explainable Artificial Intelligence (XAI).* IEEE Access, 2025. **13**: p. 41111–41140.
12. Comito, C. and C. Pizzuti, *Artificial intelligence for forecasting and diagnosing COVID-19 pandemic: A focused review.* Artificial Intelligence in Medicine, 2022. **128**: p. 102286.
13. Mohapatra, N., K. Shreya, and A. Chinmay. *Optimization of the Random Forest Algorithm*. in *Advances in Data Science and Management*. 2020. Singapore: Springer Singapore.

14. Belyadi, H. and A. Haghighat, *Chapter 5 - Supervised learning*, in *Machine Learning Guide for Oil and Gas Using Python*. 2021, Gulf Professional Publishing. p. 169–295.
15. Tonin, B.S.H., et al., *Predictive modeling of composite fracture toughness using machine learning.* Dental Materials, 2026.