Interpretable Machine Learning for Predicting the Fatigue Strength of Steel: Influence of Composition and Processing Parameters

Soheila Kookalani[1], Gozde Basak Ozturk[2], Hamidreza Alavi[1], Erika Parn[1], and Ioannis Brilakis[1]

[1]Department of Engineering, University of Cambridge, Cambridge, UK
[2]Department of Civil Engineering, Aydin Adnan Menderes University, Aydin, Turkey
sk2268@cam.ac.uk, gbozturk@edu.edu.tr, sa2194@cam.ac.uk, eap47@cam.ac.uk, ib340@cam.ac.uk

Abstract –

The prediction of fatigue strength in steel is critical for the design and analysis of structural components, given the high costs and time associated with fatigue testing and the severe consequences of fatigue failures. This study explores the application of the CatBoost machine learning algorithm to predict the fatigue strength of steel based on chemical composition, processing parameters, and mechanical properties. The model's interpretability is enhanced by integrating Shapley additive explanations, providing insights into the contributions of key features such as tempering temperature (TT), chromium content (Cr), and molybdenum content (Mo). The proposed framework achieves high predictive accuracy, with an R^2 of 0.952 and an RMSE of 31.625. This study fosters trust and utility in safety-critical applications by addressing the limitations of black-box models. The results underscore the potential of interpretable machine learning in advancing fatigue strength prediction methodologies and informing structural and construction engineering practices, while Enhancing the SHAP-based feature importance analysis to refine the selection of key predictors, potentially simplifying the model while maintaining accuracy.

Keywords –

Fatigue strength; Steel; Construction; Machine learning; Shapely additive explanations; CatBoost; Predictive modeling; Regression; Data interpretability.

1 Introduction

Accurate prediction of the fatigue strength of steels is of particular significance due to the extremely high costs and time requirements of fatigue testing, as well as the often-debilitating consequences of fatigue failures [1], [2]. Fatigue strength is the most fundamental data required for the design and failure analysis of structural components [3], [4], [5]. It is reported that fatigue accounts for over 90% of all mechanical failures in structural components [6], underscoring the critical need for reliable fatigue life prediction methods.

Fatigue failure is a complex phenomenon influenced by a multitude of variables, including material composition, microstructure, processing parameters, and operational conditions [7]. The intricate interactions among these variables pose significant challenges for conventional modeling and analytical approaches, which often fail to capture the non-linear relationships and hidden dependencies inherent in fatigue behavior. Consequently, a robust and reliable predictive framework is essential for advancing understanding and improving the safety and performance of steel components.

In recent years, machine learning (ML) algorithms have demonstrated significant potential in solving complex predictive tasks in structural engineering [8], [9], [10], [11]. Wang et al. [12] provided a review of fatigue life prediction methods, highlighting the limitations of purely data-driven approaches. Zhan et al. [13] presented a ML-based approach for predicting the fatigue life of additively manufactured stainless steel 316L. However, the application of these algorithms to fatigue strength prediction has been limited, particularly in terms of interpretability. Most ML models function as "black boxes," providing accurate predictions but little insight into the decision-making processes or the influence of individual variables. This lack of transparency hampers the adoption of ML techniques in safety-critical domains like fatigue prediction.

This study seeks to address this challenge by investigating the following research question: Can an interpretable ML framework effectively predict the fatigue strength of steel while providing actionable insights into the contributions of individual features, such as chemical composition and processing parameters? To achieve this, it leverages the CatBoost ML algorithm for predicting the fatigue strength of steels and employing

Shapley additive explanations (SHAP) to enhance model interpretability. Unlike alternative interpretability techniques such as Local Interpretable Model-agnostic Explanations (LIME) or attention mechanisms in deep learning, SHAP offers a global and local understanding of feature contributions while maintaining consistency and theoretical robustness in feature attribution. This research seeks to unravel the "black box" nature of the predictive model, providing actionable insights into the roles and relative importance of individual input variables by focusing on composition and processing parameters. The proposed approach not only improves the accuracy of fatigue strength predictions but also enhances understanding of the underlying mechanisms, bridging the gap between advanced computational tools and their practical applications in material design and construction engineering.

Through this work, we aim to contribute to the advancement of fatigue strength prediction methodologies and demonstrate the value of interpretable ML in tackling complex engineering problems, ultimately fostering greater trust and utility in data-driven predictive frameworks.

2 Shapley additive explanations

SHAP is a powerful interpretability technique based on cooperative game theory, designed to provide a transparent understanding of ML model predictions [14]. Originating from the Shapley value, which was developed to fairly distribute the gains of a coalition among its participants, SHAP translates this concept into a robust mathematical framework for evaluating the contribution of each feature to a model's output. One of the defining strengths of SHAP lies in its ability to offer both global and local interpretability. On a global scale, SHAP analyzes feature importance across the entire dataset, revealing overarching patterns and identifying the key drivers influencing the predictions. This enables stakeholders to understand the broader trends and relationships captured by the model. On the local level, SHAP provides detailed insights into individual predictions, quantifying the specific contribution of each feature to a given outcome.

Impact of SHAP is enhanced by its visualization tools, which facilitate intuitive exploration of model behavior. Summary Plots provide a bird's-eye view of feature importance across the dataset. Summary plots reveal the most influential features and their directional impact (positive or negative) on the predictions by aggregating SHAP values for all features and instances. The color-coded representation (often based on feature values) enhances the clarity of this analysis. Dependence plots illustrate the relationship between a specific feature and the predicted outcomes, capturing non-linear effects and feature interactions. These visualizations help uncover intricate dependencies, shedding light on how features interact to influence predictions.

While SHAP offers a theoretically sound and practically valuable approach, it comes with some computational challenges. Computing exact Shapley values can be resource-intensive, especially for high-dimensional datasets or complex models. Approximation techniques, such as those implemented in SHAP libraries, mitigate this issue but may still pose scalability concerns for extremely large datasets. Alternative interpretability methods, such as LIME, provide faster but less stable feature attributions, while inherently interpretable models like decision trees and Explainable Boosting Machines (EBMs) offer transparency at the cost of predictive accuracy.

Despite these challenges, SHAP remains a cornerstone of model interpretability. Its rigorous foundation, adaptability to various ML models, and rich visualization toolkit empower users to make informed decisions, enhance model trust, and align ML applications with ethical and regulatory standards.

3 Numerical examples

In addressing the inherent black-box nature of ML models, particularly in the context of predicting the fatigue strength of steel, this study adopts an interpretable ML approach. The lack of transparency in black-box models often leads to reduced trust and understanding, posing challenges to their broader acceptance and applicability in critical engineering fields.

The Fatigue Dataset for Steel, obtained from the National Institute of Material Science (NIMS) MatNavi [15], serves as the basis for this study. This dataset is one of the largest globally, offering detailed information on chemical composition, upstream mill product features, and subsequent processing parameters. It includes data for carbon, low alloy, carburizing, and spring steels, with the target property being fatigue strength, derived from rotating bending fatigue tests conducted under room temperature conditions.

The dataset comprises 437 instances with 25 features related to composition, processing parameters, and mechanical properties. The features are categorized as follows:

- Chemical Composition: %C, %Si, %Mn, %P, %S, %Ni, %Cr, %Cu, %Mo (all in weight percentage).
- Upstream Processing Details: Ingot size, reduction ratio, and non-metallic inclusion characteristics.
- Heat Treatment Conditions: Parameters such as temperature, time, and other process-specific conditions for normalizing, through-hardening, carburizing-quenching, and tempering processes.
- Mechanical Properties: Yield strength (YS),

ultimate tensile strength (UTS), elongation (%EL), reduction of area (%RA), hardness, Charpy impact value (J/cm²), and fatigue strength.

Out of the 437 data instances, the breakdown is as follows: 371 carbon and low-alloy steels, 48 carburizing steels, 18 spring steels. These data instances represent various heats within each grade of steel and different processing conditions, providing a diverse and comprehensive dataset for predictive modeling.

The target property for this study is the fatigue strength of steel, while the features include a combination of chemical composition, processing parameters, and mechanical properties. This comprehensive selection enables the model to capture the complex interactions influencing fatigue behavior. Using the dataset, the study constructs predictive models with the CatBoost ML algorithm and applies SHAP to interpret the model's predictions. The integration of SHAP enhances transparency by identifying the contribution of each feature to the predicted fatigue strength, addressing the black-box nature of ML models, and facilitating actionable insights.

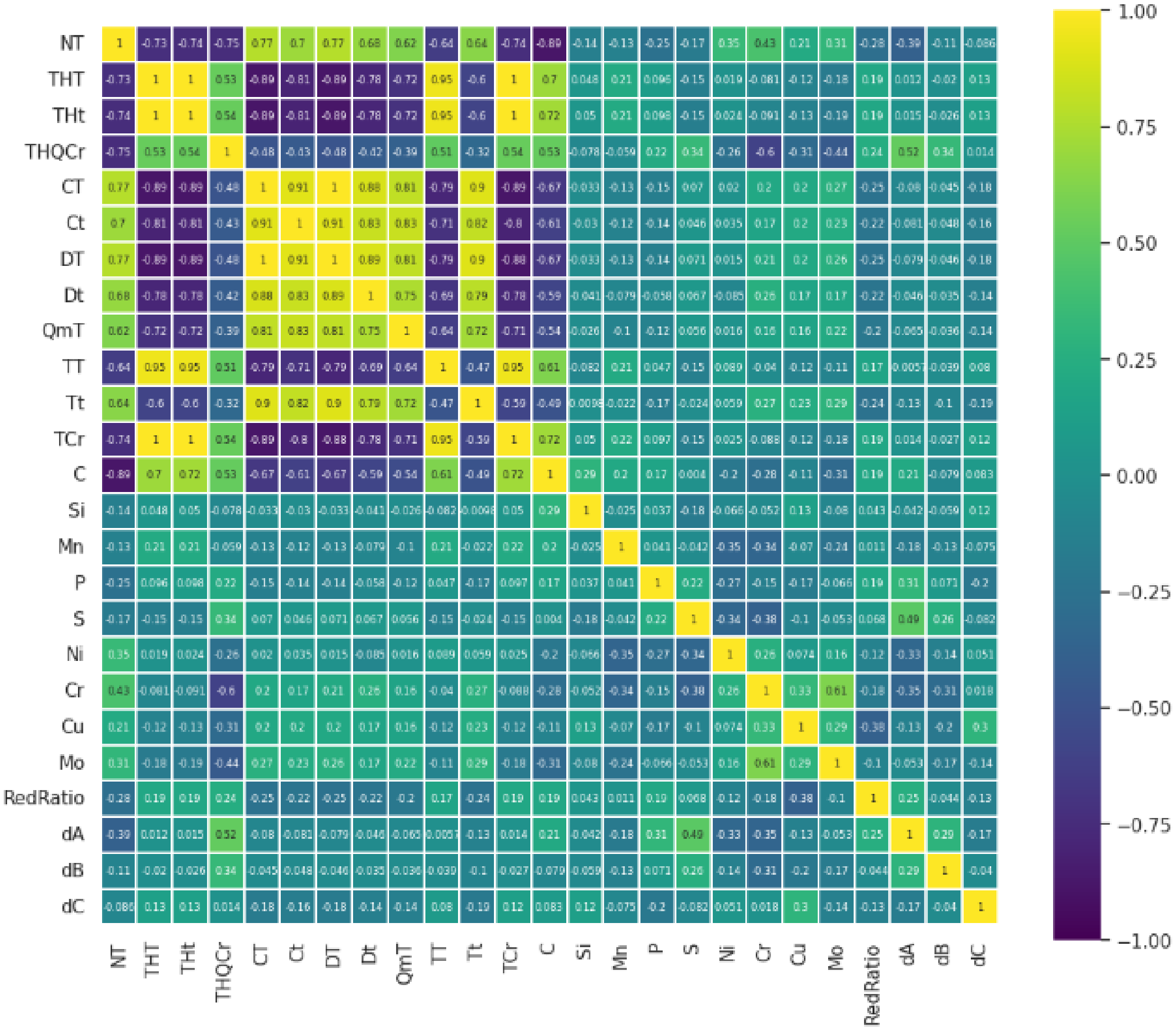


Figure 1. Correlation matrix

Details of the 25 features are provided in Table 1, offering a complete overview of the variables utilized in the model and their relevance to fatigue strength prediction. This numerical example demonstrates the practical application of an interpretable ML framework.

Figure 1 presents a correlation matrix illustrating the pairwise correlations between various numerical features, represented by the rows and columns of the matrix. The color scale on the right indicates the range of correlations, from -1 (dark purple) to 1 (yellow). Positive correlations are depicted in green to yellow, while negative correlations appear in blue to purple. The diagonal values are all 1, as they represent the correlation of each feature with itself. Certain feature pairs exhibit strong positive correlations (close to 1, highlighted in yellow), including THt and THCr (~0.89), TT and THt (0.95), and Ct and DT (0.91). These strong correlations suggest a linear relationship between these variables. Conversely, some features show strong negative correlations (close to -1, dark purple), such as NT and THt (-0.74) and CT and QmT (-0.67). In contrast, some feature pairs have correlations close to 0 (e.g., NT and dC), suggesting no linear relationship. Highly correlated features could introduce multicollinearity issues, potentially necessitating their removal or combination. Conversely, features with low or no correlation to others may warrant further investigation.

Table 1 Data features

Abbreviation	Details
C	% Carbon
Si	% Silicon
Mn	% Manganese
P	% Phosphorus
S	% Sulphur
Ni	% Nicke
Cr	% Chromium
Cu	% Copper
Mo	% Molybdenum
NT	Normalizing Temperature
THT	Through Hardening Temperature
THt	Through Hardening Time
THQCr	Cooling Rate for Through Hardening
CT	Carburization Temperature
Ct	Carburization Time
DT	Diffusion Temperature
Dt	Diffusion time
QmT	Quenching Media Temperature (for Carburization)
TT	Tempering Temperature
Tt	Tempering Time
TCr	Cooling Rate for Tempering
RedRatio	Reduction Ratio (Ingot to Bar)
dA	Area Proportion of Inclusions Deformed by Plastic Work
dB	Area Proportion of Inclusions Occurring in Discontinuous Array
dC	Area Proportion of Isolated Inclusions
Fatigue	Rotating Bending Fatigue Strength (10^7 Cycles)

3.1 Regression model

A comparative study is conducted on several ML algorithms, including Decision Tree (DT), LightGBM, and CatBoost, to determine the most accurate one. Table 2 presents the R^2 and RMSE values for each algorithm. Hyperparameter tuning is performed using a grid search approach combined with 10-fold cross-validation to prevent overfitting. The hyperparameter values that yield the best performance are then selected for the ML models.

Table 2 Model performance

ML model	R^2	RMSE
DT	0.896	39.257
LightGBM	0.893	54.209
CatBoost	0.952	31.652

As a result of the comparative study, the CatBoost method achieved the highest accuracy with the optimal hyperparameter values as follows: Depth = 5, Iterations = 1000, and Learning Rate = 0.05. These optimal values were selected from the following ranges: Depth {5, 6, 7}; Iterations {500, 1000, 1500}; and Learning Rate {0.01, 0.03, 0.05}. Figure 2 presents the regression plot of the CatBoost model, which confirms its high accuracy. The data points generally align along a diagonal line, indicating that the predicted values closely match the actual values. Minor deviations from the diagonal suggest potential data quality issues, overfitting, or specific conditions affecting predictions. A strong correlation between the actual and predicted data is observed. The 10-fold cross-validation results in average R^2 and RMSE values of 0.952 and 31.652, respectively. Since the fatigue strength values in the dataset range widely from 225 MPa to 1190 MPa, an RMSE of 31.625 is relatively small, indicating good predictive performance. In the subsequent subsection, the CatBoost model is further analyzed using an interpretable method.

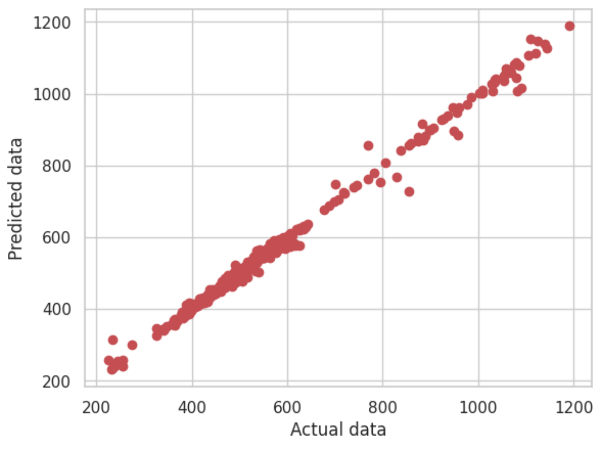


Figure 2. Regression plot

3.2 Interpretable method

Figure 3 represents the feature importance in CatBoost model, where each feature's contribution (relative importance) is measured and ranked. TT is the most significant feature, dominating the importance ranking. It dominates the analysis compared to other features. Tt, QmT, and Cr are the next most important features. There is a gradual decline in importance as we move down the list, with features like THt and RedRatio contributing minimal importance. The most important features have the largest impact on the performance of the model. Lower-ranked features contribute relatively little and might be considered for potential feature reduction or pruning in future analyses.

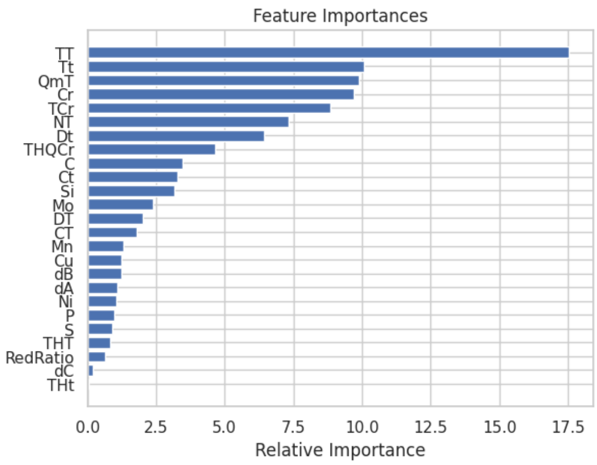


Figure 3. CatBoost importance factor

Figure 4 shows a SHAP summary plot. The features TT, Cr, and Mo have the largest SHAP values, meaning they have the most significant impact on the model's predictions. Their impact spans a wide range of positive and negative values. Features like Cr and Mo show a color gradient, meaning that low feature values (blue) and high feature values (red) influence predictions differently. For instance, high values of Cr (red) push predictions to the right (positive SHAP), while low values (blue) may push them to the left. Features lower down the list, like RedRatio, dA, dB, and P, have a much smaller spread of SHAP values. The varying spread of SHAP values (horizontal spread) reflects how much variance each feature causes in the predictions.

Figure 5 shows a SHAP summary plot that shows the importance of various features. TT has the highest importance in the model, contributing the most to the predictions. Cr follows closely, also playing a major role. Mo, QmT, Dt, and C show moderate impact compared to the top two. Features such as RedRatio, dA, dB, and P have very low SHAP values. The plot shows a clear skew towards the top features (TT, Cr), meaning the model's predictions rely heavily on just a few variables. Identifying the top contributing features (TT and Cr) helps focus on those key variables for further analysis or model optimization. Features with minimal contributions could potentially be excluded to simplify the model.

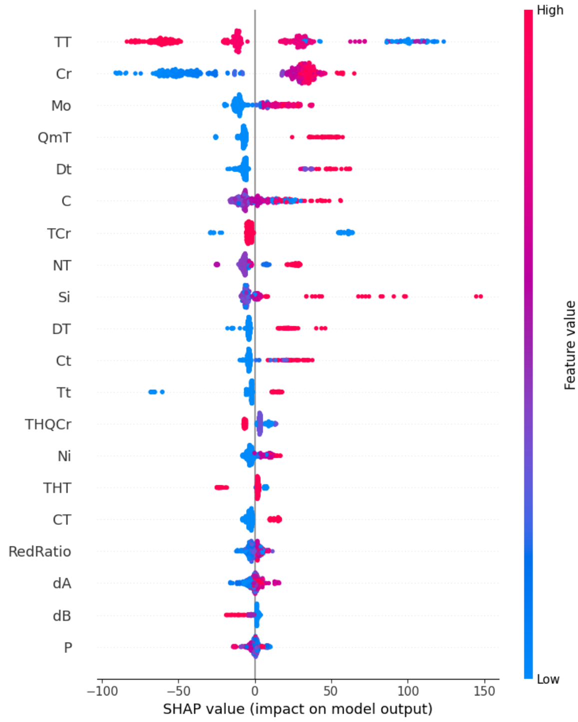


Figure 4. SHAP plots: Shapely value

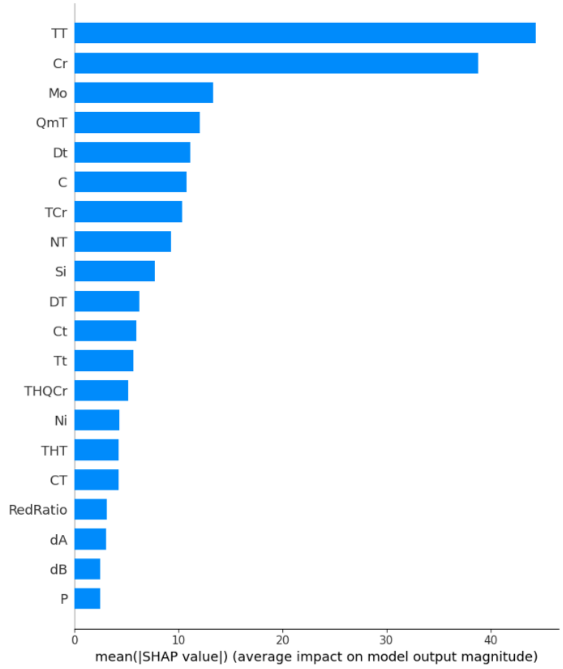
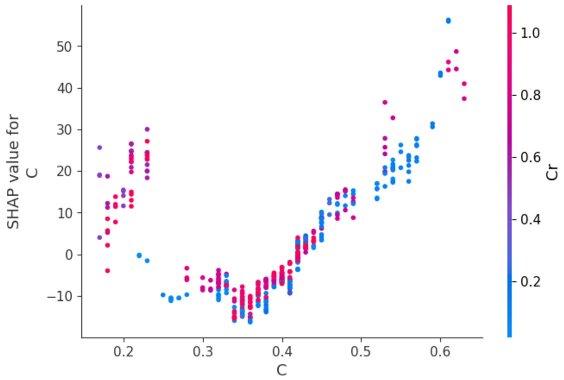
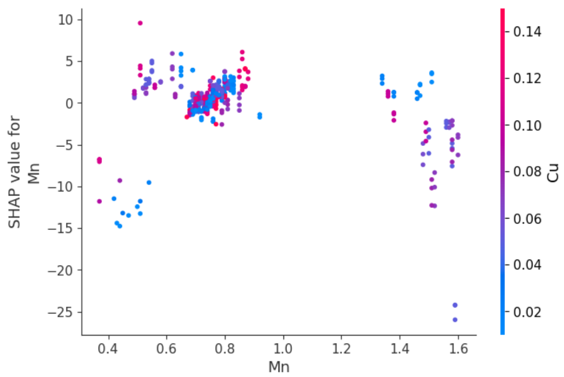


Figure 5. SHAP plots: Global importance factor.

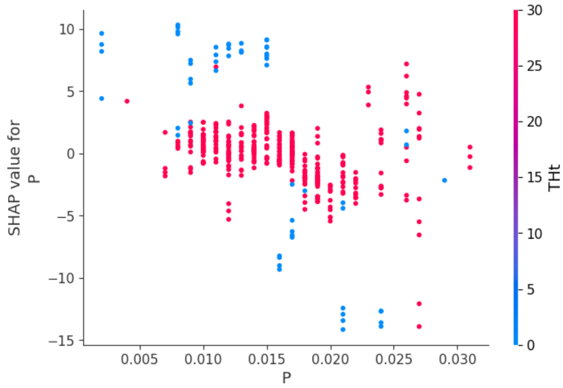
Figure 6 illustrates SHAP value plots that explain the contributions of various features to the model's predictions. Figure 6a shows that SHAP values increase non-linearly with rising "C" values, exhibiting a U-shaped pattern. Figure 6b highlights that the "Mn" feature forms two clusters with distinct SHAP patterns. Figure 6c demonstrates that SHAP values for "P" remain slightly positive for lower "P" values but become more dispersed as "P" increases. Figure 6d reveals a clear positive relationship between "S" and its SHAP values. Figure 6e indicates that SHAP values rise with "Ni" but are clustered into groups. Figure 6f shows that SHAP values increase with "Cr" in a strong positive linear trend. Figure 6g illustrates that SHAP values for "Cu" are mostly scattered around zero, with a slight positive correlation. Figure 6h demonstrates a positive relationship between SHAP values and "Mo" as "Mo" increases. Figure 6i shows that SHAP values fluctuate widely for lower "RedRatio" values but stabilize as "RedRatio" increases. Figure 6j indicates a slight upward trend in SHAP values as "dA" increases. Higher "Cr" appears to significantly influence the SHAP range. Overall, features such as "C," "Cr," "Mo," and "Ni" exhibit clear trends with SHAP values, while others, like "Cu" and "P," show more scattered distributions. These SHAP patterns suggest that specific features strongly contribute to predictions under particular conditions or value ranges.



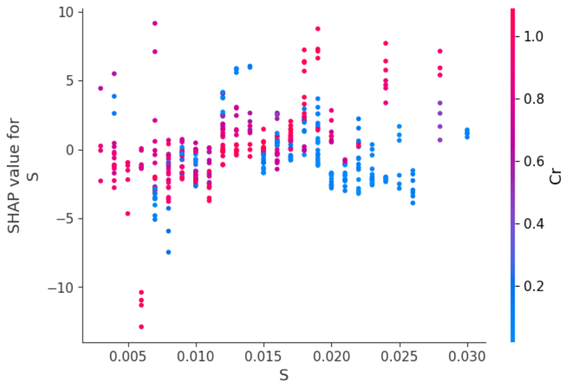
(a)



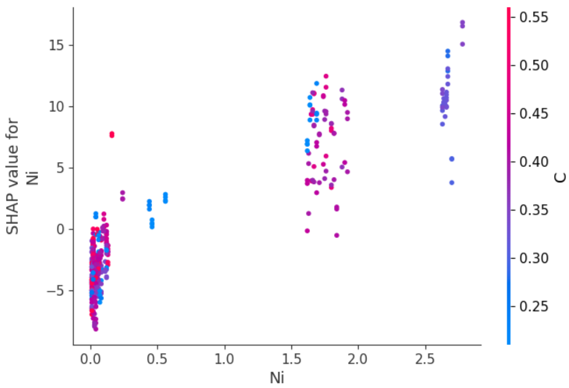
(b)



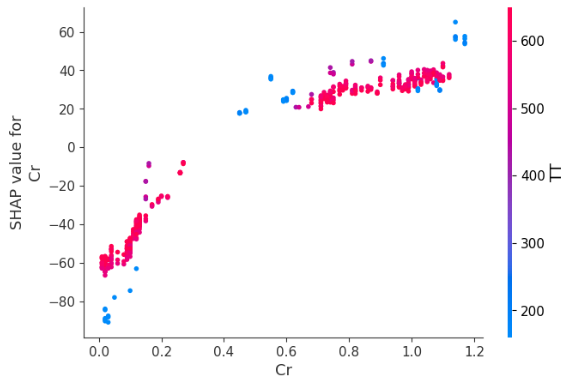
(c)



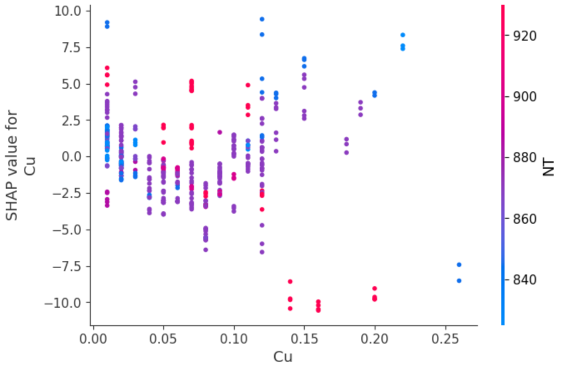
(d)



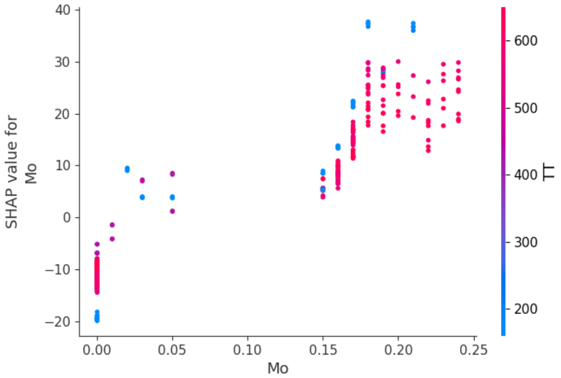
(e)



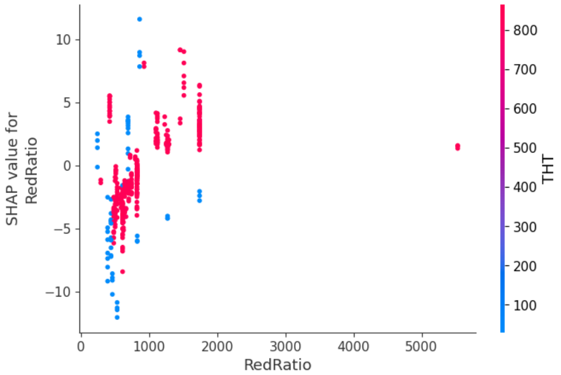
(f)



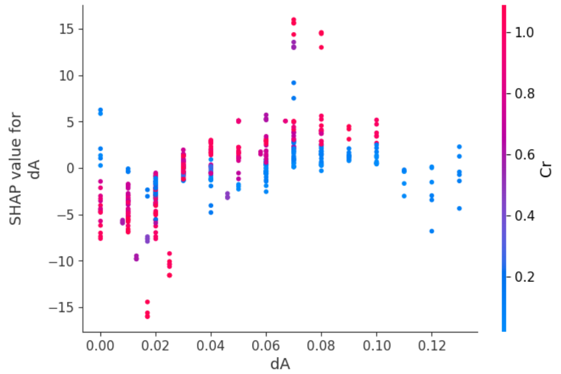
(g)



(h)



(i)



(j)

Figure 6. SHAP partial dependence plots.

4 Conclusions

This study demonstrates the significant potential of ML techniques, particularly the CatBoost algorithm, in accurately predicting the fatigue strength of steel while addressing the interpretability challenges commonly associated with black-box models. This research not only enhances predictive accuracy but also provides meaningful insights into the underlying relationships driving fatigue strength by incorporating a diverse dataset encompassing chemical composition, processing parameters, and mechanical properties. The use of SHAP further enables the decomposition of model predictions, offering a dual perspective on global trends and instance-specific behavior.

The analysis revealed that features such as tempering temperature (TT), chromium content (Cr), and molybdenum content (Mo) are the most influential in determining fatigue strength, with tempering temperature exhibiting the highest importance across predictions. The SHAP visualizations illustrated the nonlinear relationships and interactions among variables, highlighting how different combinations of features contribute to the overall predictions. For instance, the study identified synergistic effects between certain alloying elements and processing conditions, underscoring the complexity of fatigue behavior in steel and the importance of considering multivariate interactions in predictive modeling.

This research emphasizes the importance of interpretability in ML applications. The integration of SHAP enhances transparency, enabling engineers to gain actionable insights into the factors influencing fatigue strength. This interpretability fosters trust in the model's outputs, making it more suitable for safety-critical domains, such as structural engineering, where decision-making must be reliable and well-informed. From a practical perspective, this framework can support engineers in optimizing steel selection and processing decisions, reducing fatigue-related failures, and enhancing the sustainability and reliability of steel structures in construction engineering.

Furthermore, the study identified features with minimal contributions to model predictions, such as the reduction ratio and certain inclusion properties, which could be candidates for exclusion in future modeling efforts. This simplification has the potential to reduce computational demands and streamline predictive frameworks without sacrificing accuracy. Future research should explore feature selection methods to reduce redundancy and improve model efficiency, alongside alternative ML algorithms for comparison.

Future research can expand upon this framework by exploring the applicability of interpretable ML techniques to other material properties and engineering challenges, such as fracture toughness or corrosion resistance. Additionally, incorporating larger and more diverse datasets, including environmental and operational factors, could further enhance the robustness and generalizability of the predictive models. The integration of additional techniques, such as explainable deep learning or hybrid models combining physics-based and data-driven approaches, may also provide deeper insights into complex structural behaviors.

Acknowledgment

The financial support provided by the EPSRC for the RESTOR project (Grant No. EP/W018705/1) and by the EU Commission for the AEGIR project (Grant No. 101079961) is gratefully acknowledged.

References

[1] G. Zuo, L. Tong, Z. Zhao, H. Wang, W. Shi, and Y. Yan, "Experimental study on fatigue strength of trapezoidal corrugated-web steel girders based on hot spot stress method," *Thin-Walled Structures*, vol. 208, p. 112814, Mar. 2025, doi: 10.1016/J.TWS.2024.112814.

[2] X. Guan, H. Guo, G. Li, D. Yang, and Y. Pan, "Fatigue behaviour of welded joints of Q690D high strength steel," *J Constr Steel Res*, vol. 223, p. 109037, Dec. 2024, doi: 10.1016/J.JCSR.2024.109037.

[3] S. He, H. Zhong, X. Huang, Y. Xu, and A. S. Mosallam, "Experimental investigation on shear fatigue behavior of perfobond strip connectors made of high strength steel and ultra-high performance concrete," *Eng Struct*, vol. 322, p. 119181, Jan. 2025, doi: 10.1016/J.ENGSTRUCT.2024.119181.

[4] F. Xu, J. He, X. Zhou, Y. Wang, Q. Yang, and Y. Huang, "Experimental investigation on low-cycle fatigue and lifetime estimation of high-strength lockbolt material ML40Cr steel," *Constr Build Mater*, vol. 449, p. 138317, Oct. 2024, doi: 10.1016/J.CONBUILDMAT.2024.138317.

[5] P. Mendes *et al.*, "High-cycle fatigue design curves of mild- and high-strength steels for offshore applications," *Structures*, vol. 67, p. 106827, Sep. 2024, doi: 10.1016/J.ISTRUC.2024.106827.

[6] P. D. Deshpande, B. P. Gautham, A. Cecen, S. Kalidindi, A. Agrawal, and A. Choudhary, "Application of Statistical and Machine Learning Techniques for Correlating Properties to Composition and Manufacturing Processes of Steels," *2 nd World Congress on Integrated Computational Materials Engineering* , pp. 155–160, Jul. 2013, doi: 10.1002/9781118767061.CH25.

[7] A. Agrawal, P. D. Deshpande, A. Cecen, G. P. Basavarsu, A. N. Choudhary, and S. R. Kalidindi, "Exploration of data science techniques to predict fatigue strength of steel from composition and processing parameters," *Integr Mater Manuf Innov*, vol. 3, no. 1, pp. 90–108, Dec. 2014, doi: 10.1186/2193-9772-3-8/FIGURES/7.

[8] S. Kookalani, B. Cheng, and S. Xiang, "Shape optimization of GFRP elastic gridshells by the weighted Lagrange ε-twin support vector machine and multi-objective particle swarm optimization algorithm considering structural weight," *Structures*, vol. 33, pp. 2066–2084, 2021, doi: 10.1016/j.istruc.2021.05.077.

[9] S. Kookalani, S. Nyunn, and S. Xiang, "Form-finding of lifting self-forming GFRP elastic gridshells based on machine learning interpretability methods," *Structural Engineering and Mechanics*, vol. 84, no. 5, pp. 605–618, 2022, doi: 10.12989/sem.2022.84.5.605.

[10] S. Kookalani, B. Cheng, and J. L. C. Torres, "Structural performance assessment of GFRP elastic gridshells by machine learning interpretability methods," *Frontiers of Structural and Civil Engineering*, vol. 16, no. 10, pp. 1249–1266, 2022, doi: 10.1007/s11709-022-0858-5.

[11] S. Kookalani *et al.*, "Trajectory of building and structural design automation from generative design towards the integration of deep generative models and optimization: A review," *Journal of Building Engineering*, vol. 97, p. 110972, Nov. 2024, doi: 10.1016/J.JOBE.2024.110972.

[12] H. Wang, B. Li, J. Gong, and F. Z. Xuan, "Machine learning-based fatigue life prediction of metal materials: Perspectives of physics-informed and data-driven hybrid methods," *Eng Fract Mech*, vol. 284, p. 109242, May 2023, doi: 10.1016/J.ENGFRACMECH.2023.109242.

[13] Z. Zhan and H. Li, "Machine learning based fatigue life prediction with effects of additive manufacturing process parameters for printed SS 316L," *Int J Fatigue*, vol. 142, p. 105941, Jan. 2021, doi: 10.1016/J.IJFATIGUE.2020.105941.

[14] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," *Adv Neural Inf Process Syst*, vol. 30, 2017.

[15] "National Institute of Materials Science", doi: http://smds.nims.go.jp/fatigue/index_en.html.