

A DATA-DRIVEN MACHINE LEARNING APPROACH FOR NUCLEAR POWER PLANT APPLICATIONS

Muhammad Waqar Nasir¹, Muhammad Zubair Sheikh^{*2}, Rehmat Bashir³, Shahrob Ghazanfar⁴,
Abdul Saboor⁵, Abdullah Safdar⁶

^{1,*2,3,4,5,6}Department of Mechanical Engineering, University of Engineering and Technology (Main Campus), Lahore
54890, Pakistan

^{*2}waqarnasir@uet.edu.pk

DOI: <https://doi.org/10.5281/zenodo.22010039>

Keywords

Stress corrosion cracking; stainless steel 304; crack growth rate; LightGBM; XGBoost; machine learning; nuclear power plants; data augmentation.

Article History

Received: 16 May 2026

Accepted: 12 June 2026

Published: 15 August 2026

Copyright @Author

Corresponding Author: *

Muhammad Waqar Nasir

Abstract

Stress corrosion cracking (SCC) is a major degradation mechanism in safety-critical components because it results from the combined action of a susceptible material, tensile stress, and a corrosive environment. This study develops a data-driven model for predicting SCC crack growth rate in stainless steel 304 using six input variables: cold work, Vickers hardness, carbon content, temperature, stress intensity factor, and yield strength. The original project compiled 90 experimental data records from published research and technical sources. Because the dataset was small for machine-learning applications, regression-based data extension with bounded random perturbations was used to generate a working dataset of 1090 records while retaining the observed trends. XGBoost was initially evaluated but produced unstable predictions on the available data. Light Gradient Boosting Machine (LightGBM) was subsequently selected because it provided lower prediction errors, faster training, and more consistent behavior. The final model used regression learning with an 80:20 data split, feature standardization, a learning rate of 0.05, 31 leaves per tree, and early stopping based on validation RMSE. Model analysis identified temperature, Vickers hardness, and yield strength as the three most influential variables, with normalized importance values of 19.47%, 17.70%, and 16.90%, respectively. The predicted crack growth rate increased with stress intensity factor and showed a particularly strong temperature dependence. The reported prediction error decreased from initial values of -753.75% and 66.87% to below $\pm 1\%$ during subsequent model training, with a minimum reported error of -0.04%. Although the results demonstrate the potential of machine learning for SCC assessment, the limited experimental database and use of synthetic data mean that further validation using independent experimental and plant-relevant data is required.

1. INTRODUCTION

Nuclear power plant components operate under combinations of high temperature, pressure, mechanical loading, and chemically active environments [1]. Under such conditions, stress

corrosion cracking can initiate and propagate in susceptible materials and may threaten structural integrity even when the applied stress is below the conventional yield strength. SCC is therefore important for life assessment, inspection

planning, and preventive maintenance of reactor pressure systems, coolant piping, steam generators, weld regions, and other safety-critical components [2].

The physical behavior of SCC is difficult to describe with a single deterministic relationship because material properties, stress state, temperature, and environmental chemistry interact nonlinearly. Traditional crack-growth relations remain useful for engineering interpretation, but their applicability is generally restricted to particular materials and environmental conditions. Data-driven machine-learning methods provide an alternative route because they can learn nonlinear input-output relationships directly from experimental observations [3].

The present work focuses on stainless steel 304 and develops a regression model for crack growth rate using six variables selected from the project literature review: cold work, Vickers hardness, carbon content, temperature, stress intensity factor K , and yield strength. The central contribution is the development and assessment of a LightGBM-based prediction framework after comparison with XGBoost[4-5].

1.1 Research problem and objectives

The research problem is the difficulty of predicting SCC crack growth accurately when several material and operating parameters act simultaneously. The study therefore aims to: (i)

identify the principal variables associated with SCC crack growth; (ii) assemble a usable experimental dataset; (iii) address the limited-data problem through controlled data extension; (iv) compare XGBoost and LightGBM; (v) evaluate the final model through prediction error and trend analysis; and (vi) interpret the predicted dependence of crack growth on the selected variables.

2. Stress Corrosion Cracking and Influencing Variables

SCC requires the simultaneous presence of a susceptible material, tensile stress, and a suitable corrosive environment. Cracking may occur through grain boundaries (intergranular SCC) or through grains (transgranular SCC), depending on the material condition and environment. In stainless steels, chloride contamination, elevated temperature, residual stress, sensitization, and local changes in crack-tip chemistry can increase susceptibility.

Stainless steel 304 is widely used because of its combination of corrosion resistance, strength, ductility, and availability. Nevertheless, it can undergo SCC under unfavorable combinations of temperature, tensile stress, and environmental chemistry. Stress concentrations at welds, notches, and surface imperfections can promote crack initiation, while the local crack-tip environment controls subsequent propagation.

2.1 Selected input variables

Variable	Engineering relevance
Cold work (%)	Plastic deformation below recrystallization temperature; increases dislocation density and residual stress.
Vickers hardness (HV)	Indicator of resistance to plastic deformation and material condition.
Carbon content (mass %)	Influences microstructure, carbide formation, and sensitization.
Temperature (°C)	Controls reaction and diffusion kinetics and strongly affects SCC severity.
Stress intensity factor, K (MPa $\sqrt{m}$)	Represents the crack-tip mechanical driving force.
Yield strength (MPa)	Characterizes resistance to plastic deformation and the mechanical state of the material.

2.2 Conventional crack-growth relationships

The project discusses empirical crack-growth relationships in which crack growth rate is expressed as a function of stress intensity factor and material/environmental constants. A threshold stress intensity factor, K_{ISCC} , is also used to describe the onset of SCC, with SCC growth expected when the applied stress intensity exceeds the relevant threshold.

$$K > K_{ISCC}$$

Temperature dependence may be represented by an Arrhenius-type relationship, while cyclic corrosion fatigue can be described using Paris-type relations. These formulations provide useful physical interpretation, but the present study uses machine learning to represent the combined nonlinear effect of multiple variables.

3. Literature Review and Research Gap

Previous SCC studies have established the importance of material condition, temperature, stress intensity, and environmental chemistry. Research on austenitic stainless steels and nickel-based alloys in nuclear environments has shown that SCC susceptibility can change markedly with temperature and water chemistry. Machine-learning approaches—including artificial neural networks, random forests, support-vector regression, Gaussian-process regression, XGBoost, and LightGBM—have increasingly been considered for complex degradation problems. Tree-based ensemble methods are particularly attractive for tabular engineering data because they can represent nonlinear relationships and interactions without requiring a predefined analytical form. XGBoost uses regularized gradient boosting and level-wise tree growth, whereas LightGBM uses histogram-based processing and leaf-wise tree growth to reduce computational cost. In the project dataset, LightGBM gave more consistent predictions and lower reported errors than XGBoost.

3.1 Research gap

A central limitation identified in the project literature is the scarcity of high-quality labeled SCC data. Plant measurements are difficult to

obtain and controlled laboratory experiments are expensive. Consequently, there is a need for predictive approaches that can extract useful information from relatively small datasets while remaining physically interpretable. The present study addresses this gap through controlled data extension followed by gradient-boosting regression.

4. Methodology

The methodology consists of five main stages: experimental data collection, data extension, preprocessing, machine-learning model development, and prediction/trend analysis.

4.1 Data collection

The project collected 90 unique observations from peer-reviewed publications and technical documents dealing with SCC in austenitic stainless steels, particularly SS 304 and SS 316. The six selected input variables were cold work, Vickers hardness, carbon content, temperature, stress intensity factor, and yield strength. The target variable was crack growth rate, da/dt , expressed in mm/s.

4.2 Data extension

Because 90 records are limited for training a nonlinear machine-learning model, the project used regression-based data extension in Microsoft Excel. Linear regression equations were obtained from the available data and applied with bounded random noise to generate additional observations. The resulting working dataset contained 1090 entries. The intention was to increase statistical variability while retaining the general relationships present in the original observations.

This step should be regarded as a data-enrichment strategy rather than a substitute for new experiments. Synthetic observations inherit the assumptions and limitations of the original data and can therefore introduce bias if they are not independently validated.

4.3 Preprocessing and data split

The six input variables were converted to numerical form, missing values were handled, and

feature scaling was performed using standard normalization. The dataset was divided into training and testing subsets using an 80:20 ratio. StandardScaler was used so that the model received inputs on a comparable numerical scale.

4.4 XGBoost model

XGBoost was first implemented as the baseline gradient-boosting model. It was selected because of its ability to model nonlinear relationships, its regularization mechanisms, and its suitability for regression problems. However, the initial model produced unstable and unpredictable outputs. The project attributed this behavior primarily to overfitting of the extended dataset and/or unsuitable parameter settings.

4.5 LightGBM model

LightGBM was subsequently adopted as the principal model. The model was configured for

regression with RMSE used as the evaluation criterion. A learning rate of 0.05 and 31 leaves per tree were reported. Early stopping was used, with training terminated when validation RMSE failed to improve for 50 iterations. The model reached convergence at approximately 1090 boosting iterations in the project implementation.

4.6 Feature importance

Feature importance was calculated from the trained LightGBM model. The normalized importance values quantify the relative contribution of the six input variables to the model's tree-splitting decisions. These values are useful for engineering interpretation but should not be interpreted as direct causal coefficients.

Feature	Normalized importance
Temperature	19.47%
Vickers hardness	17.70%
Yield strength	16.90%
Cold work	15.84%
Stress intensity factor, K	15.36%
Carbon content	14.73%

5. Results

5.1 Model selection

The project found that XGBoost was competitive but less stable on the available dataset. LightGBM

produced lower reported MAE and RMSE and a higher R^2 than XGBoost, while also requiring less computational effort. On this basis, LightGBM was selected as the final prediction model.

Criterion	XGBoost	LightGBM
Prediction behavior	Less stable on the project dataset	More consistent
Training efficiency	Higher computational demand	Faster and more memory-efficient
Reported error	Higher	Lower
Final selection	Baseline/comparison model	Selected model

5.2 Feature-importance results

Temperature was the most influential variable in the trained model at 19.47%, followed by Vickers hardness at 17.70% and yield strength at 16.90%. Cold work, K, and carbon content contributed 15.84%, 15.36%, and 14.73%, respectively. The relatively close values indicate that the model does

not rely on a single input variable; rather, it represents SCC as a multivariable phenomenon.

5.3 Effect of carbon content

With cold work, Vickers hardness, and yield strength held constant in the project analysis, the predicted crack growth rate initially increased

slightly as carbon content rose from 0.001% to 0.0045%, changing from approximately 3.06×10^{-7} to 3.12×10^{-7} mm/s. At a higher carbon level of about 0.047%, the calculated rate decreased to approximately 1.45×10^{-7} mm/s. The project interpreted this nonlinear response as possible threshold behavior associated with microstructural or passive-film effects. Because the result is model-derived and the available dataset is limited, the trend should be verified experimentally before drawing a general materials conclusion.

5.4 Effect of stress intensity factor

The model predicted a steady increase in crack growth rate as K increased. At a constant carbon content of 0.012%, the rate increased from approximately 3.02×10^{-7} mm/s at $K = 13$ MPa $\sqrt{m}$ to about 3.4×10^{-7} mm/s near $K = 50$ MPa $\sqrt{m}$. This trend is consistent with the role of K as the mechanical driving force for crack extension.

5.5 Effect of temperature

Temperature produced the strongest trend among the variables examined in the project. At a carbon content of 0.047%, the predicted crack growth rate increased from approximately 1.46×10^{-7} mm/s at 250°C to about 9.4×10^{-7} mm/s near

800°C. The project associated this behavior with enhanced diffusion and corrosion kinetics at higher temperature. A further increase from approximately 394°C to 414°C was associated with an increase from about 2.54×10^{-7} to 3.09×10^{-7} mm/s in one trend analysis.

5.6 Effect of hardness and yield strength

The project also evaluated crack-growth trends against Vickers hardness and yield strength. Both variables were among the most influential features according to the LightGBM importance analysis. Higher hardness and yield strength alter the material's resistance to deformation and the local stress state, and their effects are coupled with temperature, stress intensity, and material condition. The model therefore indicates that these properties should be considered together rather than as isolated SCC drivers.

5.7 Prediction accuracy and convergence

The initial predictions showed large errors, including reported percentage errors of -753.75% and 66.87%. After additional model training and adjustment, subsequent predictions reached errors below $\pm 1\%$, with a minimum reported error of -0.04%. The actual-versus-predicted plots in the project showed substantially closer agreement after training than in the initial test.

Stage	Reported prediction behavior
Initial prediction	Very large errors; examples of -753.75% and 66.87%
After training	Errors reduced substantially
Later predictions	Less than $\pm 1\%$ reported error
Lowest reported error	-0.04%

6. Discussion

The results demonstrate the feasibility of using gradient-boosting machine learning to model SCC crack growth from a small collection of experimental observations. The most important finding is not simply the low final error, but the ability of the model to capture nonlinear interactions among material, mechanical, and thermal variables.

Temperature emerged as the highest-ranked feature and also showed the strongest individual trend in the reported sensitivity analysis. This is physically meaningful because temperature can accelerate chemical reactions, diffusion, and corrosion processes at the crack tip. The importance of K is also consistent with fracture mechanics because increasing K increases the local crack-tip driving force.

The carbon-content result illustrates the advantage and limitation of data-driven modeling. The model can represent non-monotonic behavior that would be difficult to capture using a simple power law. However, when the underlying dataset is small and partially synthetic, a nonlinear trend may reflect data distribution rather than a universal physical mechanism. Consequently, machine-learning predictions should be used as decision-support information and should remain linked to experimental evidence and established fracture-mechanics principles.

The feature-importance results show a relatively distributed contribution among all six variables. Temperature, hardness, and yield strength together account for approximately 54.07% of the normalized importance, while cold work, K, and carbon content account for the remaining approximately 45.93%. This distribution supports the view that SCC is a coupled material-mechanical-environmental problem rather than a single-variable phenomenon.

6.1 Engineering implications

A trained predictive model can potentially support preliminary screening of operating conditions, prioritization of inspection locations, and maintenance planning. For example, conditions associated with elevated temperature, high material hardness, high yield strength, or high K could be flagged for more detailed engineering assessment. However, the model should not replace safety codes, fracture-mechanics assessments, non-destructive examination, or plant-specific integrity evaluations.

6.2 Limitations

The main limitation is data quantity and quality. The original experimental database is small, and the expanded dataset is generated synthetically. The paper therefore cannot claim that the reported accuracy represents independent real-world prediction performance. A second limitation is the restricted feature set. Important environmental variables such as pH, chloride concentration, dissolved oxygen, and coolant flow rate were discussed in the thesis but were not included as model inputs. A third limitation is the

need for independent validation under plant-relevant conditions.

7. Conclusions and Future Work

This study developed a LightGBM-based framework for predicting SCC crack growth rate in stainless steel 304 using six input variables: cold work, Vickers hardness, carbon content, temperature, stress intensity factor, and yield strength. Based on the project results, LightGBM was more stable and accurate than the initially tested XGBoost model on the available dataset.

- Temperature was the highest-ranked model feature (19.47%), followed by Vickers hardness (17.70%) and yield strength (16.90%).
- Stress intensity factor showed a positive relationship with predicted crack growth rate, increasing from about 3.02×10^{-7} to 3.4×10^{-7} mm/s as K increased from 13 to approximately 50 MPa $\sqrt{m}$ in the reported analysis.
- At fixed carbon content of 0.047%, the predicted crack growth rate increased from about 1.46×10^{-7} mm/s at 250°C to about 9.4×10^{-7} mm/s near 800°C.
- The model's reported error improved from initial extreme values, including -753.75% and 66.87%, to below $\pm 1\%$ during subsequent training, with a minimum reported error of -0.04%.
- The results demonstrate the potential of machine learning as a supplementary tool for SCC screening, life assessment, and maintenance planning.

Future work should expand the experimental database and include environmental variables such as pH, chloride concentration, dissolved oxygen, and flow rate. Independent experimental validation is particularly important because synthetic data were used to enlarge the training dataset. Further studies may compare LightGBM with neural networks and other ensemble methods, apply explainable-AI techniques such as SHAP, and integrate validated predictions with real-time structural-health-monitoring data.

REFERENCES

- Kain, V. (2011). Stress corrosion cracking (SCC) in stainless steels. In Stress Corrosion Cracking, Elsevier, pp. 199–244.
- Andresen, P. I., & Mom, M. M. (2007). SCC of stainless steels and Ni alloys in high temperature water. NACE CORROSION.
- Terachi, T., et al. (2012). SCC growth behaviors of austenitic stainless steels in simulated PWR primary water. Journal of Nuclear Materials, 426(1–3), 59–70.
4. Bryan, C.R. and D. Enos, *Summary of available data for estimating chloride-induced SCC crack growth rates for 304/316 stainless steel*. 2016, Sandia National Lab.(SNL-NM), Albuquerque, NM (United States).
5. Abubakar, S.A., S. Mori, and J. Sumner, *A review of factors affecting SCC initiation and propagation in pipeline carbon steels*. Metals, 2022. 12(8): p. 1397

