Evaluation of Predictive Performance Based on Machine Learning Algorithm Using Ruptured Data of Zircaloy Cladding

Jungju Lee[a], Changhwan Shin[b], Sung-Uk Lee[b], Hyochan Kim[b], Byoungjae Kim[a*]
[a]*Mechanical Engr. Program, The Chungnam National Univ., 99, Daehak-ro, Yuseong-gu, Daejeon, KOREA*
[b]*LWR Fuel Technology Research Division, KAERI, 111, Daedeok-daero 989 Beon-gil, Yuseong-gu, Daejeon, KOREA*
[*]*Corresponding author: bjkim@cnu.ac.kr*

Keywords : Zircaloy cladding, machine learning, regression model

1. Introduction

Machine learning, a branch of artificial intelligence, enables computers to learn patterns and rules from data without explicit programming. This technology can be effectively utilized for the prediction and analysis of complex systems. In a nuclear reactor, a loss-of-coolant accident (LOCA) can occur when the coolant pressure drops below the internal fuel rod gas pressure, causing the cladding to expand and potentially rupture under certain conditions. Such accidents can significantly impact reactor safety, making accurate prediction and preventive measures crucial. This study applied a machine learning regression model to predict cladding rupture in LOCA scenarios using data from NUREG-0630 [1].

2. Rupture prediction methodology

2.1 Data analysis

Before applying machine learning, data preprocessing is required. Missing values and outliers were identified and removed from the experimental data. The total data consists of 216, of which 162, or 75%, were used as training data, and the remaining 54, or 25%, were used as evaluation data. The independent variables of the experimental data are heating rate and pressure, and the dependent variables are burst temperature, strain, and burst stress. Detailed information about the data is presented in Table I.

Table I: Experimental data of NUREG-0630

	Range	Average
Ramp Rate [K/s]	0 - 44	20.8
Rupture Pressure [MPa]	0.5 - 19.1	6.8
Rupture Temperature [K]	961 - 1553	1136.8
Burst Strain [%]	8 - 116	44.8
Burst Stress [MPa]	4.7 - 155.1	52.1

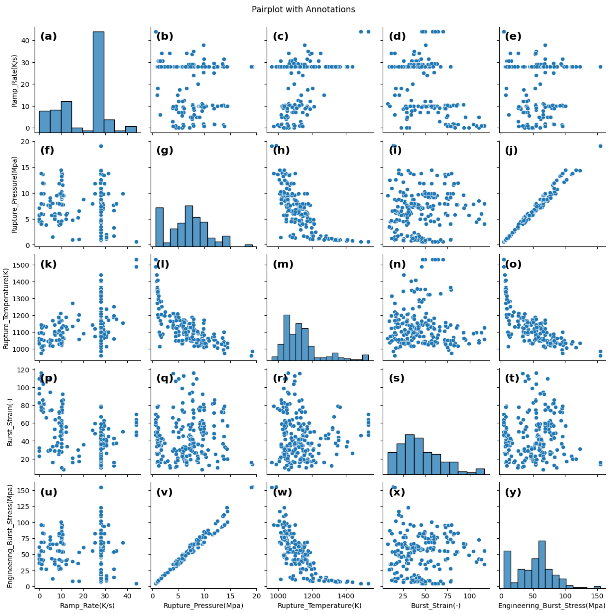

Fig. 1. Burst test data of Zircaloy cladding

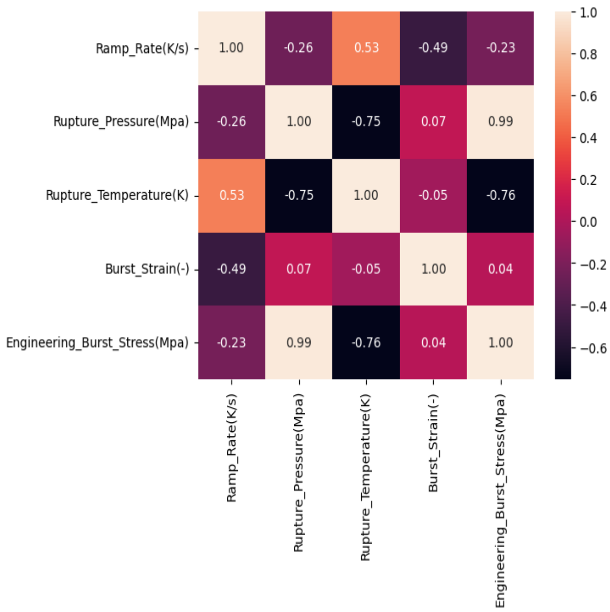

Fig. 2. Correlation coefficient between variables

The distribution of data and the relationship between variables are shown in Fig. 1, and the correlation coefficients between variables are shown in Fig. 2. A correlation coefficient closer to 1 indicates a strong positive correlation, a closer to -1 indicates a strong negative correlation, and a closer to 0 indicates no linear correlation. The correlation coefficient between the independent variable, pressure, and the dependent variable, rupture temperature is -0.75, showing an inverse proportional tendency in (l) of Fig. 1. On the other hand, the correlation coefficient between pressure and rupture stress is 0.99, showing a linear proportional tendency in (v) of Fig. 1. The strain data has no clear trend and is scattered throughout.

2.2 Machine learning algorithm

XGBoost [3] was used for regression analysis. XGBoost (eXtreme Gradient Boosting) is based on the gradient boosting algorithm that sequentially trains multiple weak learners (decision trees) and minimizes errors through gradient descent, and has improved GBM (Gradient Boosting Machine) to prevent overfitting and have efficient parallel processing functions. The model was trained with heating rate and pressure as independent variables and rupture temperature, strain, and rupture stress as dependent variables. In order to maximize model performance, the Bayesian Optimization package [4] was used to optimize the hyperparameters of XGBoost, including learning_rate (the degree of error correction of the tree), max_depth (the maximum depth of the tree), subsample (the ratio of data samples that the tree will learn), and n_estimators (the number of trees).

3. Rupture prediction results

The prediction results of the cladding failure model using XGBoost are shown in Fig. 3. The predicted values for burst temperature and burst stress are similar to the actual values; however, the difference between the predicted and actual values for strain is relatively large. The model performance was evaluated using MAPE (Mean Absolute Percentage Error). MAPE represents the absolute error between predicted and actual values as a percentage and calculates the average of these values to assess prediction accuracy. The model evaluation results are presented in Table II. For the test data, the error ranges are 2.2% and 6.3% for burst temperature and burst stress, respectively, while the strain shows a relatively large error range of 25.6%.

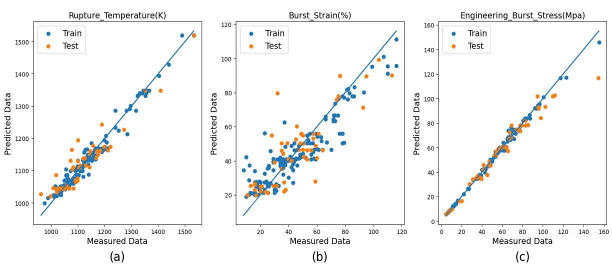

Fig. 3. Evaluation by XGBoost algorithm

Table II: XGBoost prediction performance for cladding rupture

Features	XGBoost	
	MAPE	
	Train	Test
RuptureTemperature(K)	1.2%	2.2%
Burst Strain(%)	25.4%	25.6%
Engineering Burst Stress(MPa)	3.3%	6.3%

4. Conclusions

This study performed a machine learning regression analysis based on experimental data to predict the failure of Zircaloy cladding under LOCA conditions. Analysis of the correlation coefficients between variables and the failure prediction results indicates that the correlation between independent and dependent variables affects prediction performance. When the absolute value of the correlation coefficient between any of the two independent variables and the dependent variable is closer to 1, the prediction performance is better; however, when it is closer to 0, the performance is poorer. Machine learning models are highly dependent on the quality and quantity of data, making data consistency and comprehensiveness crucial. The data used in this study were collected from multiple experiments rather than a single experiment, leading to differences in experimental conditions. Since these differences in experimental conditions may have affected the correlations between variables, analyzing these effects is necessary. Future research should aim to improve prediction performance by acquiring additional data under the same experimental conditions to enhance performance in specific scenarios or by collecting more data that covers a variety of experimental conditions to improve the model's generalization ability and reduce uncertainty.

Acknowledgements

The preparation of this paper was supported by the Ministry of Science and ICT, Republic of Korea (RS-2022-00144002)

REFERENCES

[1] D.A Powers and R.O. Meyer, Cladding Swelling and Rupture Models for LOCA Analysis, NUREG-0630, 1980.
[2] C. Shin, D. Kim, and H. Kim, Assessment on Prediction Performance of Machine Learning Algorithms for High Temperature Oxidation Data of Zircaloy Cladding, KAREI/TR-8639/2021, 2021.
[3] T. Chen and C. Guestrin, XGBoost: A scalable Tree Boosting System, In 22nd SIGKDD Conference on Knowledge Discovery and Data Mining, 2016.
[4] https://github.com/fmfn/BayesianOptimization
[5] Aurelien Geron, Hands-on Machine Learning with Scikit-Learn, Keras & TensorFlow, OReilly Media, 2019.