IJE TRANSACTIONS A: Basics Vol. 37 No. 07, (July 2024) 1296-1305

International Journal of Engineering

Journal Homepage: www.ije.ir

Effect of Training Data Ratio and Normalizing on Fatigue Lifetime Prediction of Aluminum Alloys with Machine Learning

M. Matin, M. Azadi*

Faculty of Mechanical Engineering, Semnan University, Semnan, Iran

PAPER INFO

Paper history:
Received 01 December 2023
Received in revised form 20 December 2023
Accepted 31 December 2023

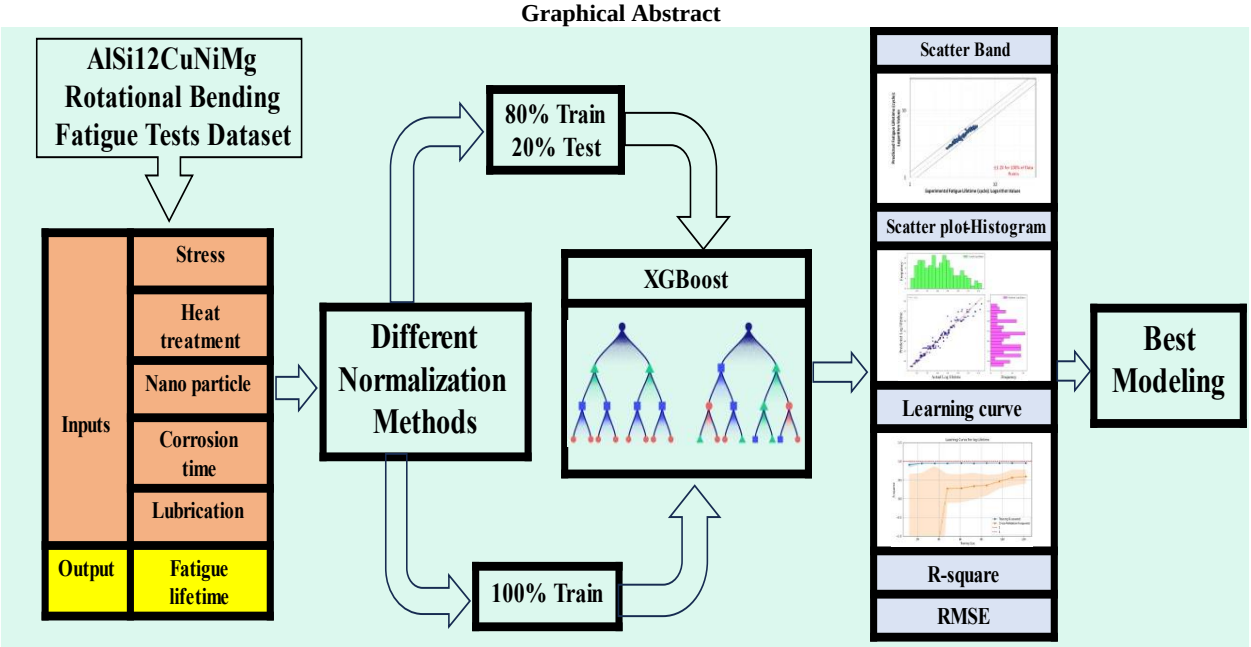
Keywords:
Machine Learning
Fatigue Lifetime
Extreme Gradient Boosting
Aluminum Alloys
Normalization Techniques
Training Data Percentage

ABSTRACT

It is critical to evaluate the estimation of the fatigue lifetimes for the piston aluminum alloys, particularly in the automotive industry. This paper investigates the effect of different normalization methods on the performance of the fatigue lifetime estimation using Extreme Gradient Boosting (XGBoost), as a supervised machine learning method. For this purpose, the dataset used in this study includes various physical and experimental inputs related to an aluminum alloy and the corresponding fatigue lifetime outputs. Furthermore, before fitting the XGBoost model, different fatigue lifetime preprocessing methods were utilized and evaluated using metrics such as Root Mean Square Error (RMSE), Determination Coefficient (R^2), and Scatter Band (SB). The results indicate that modeling fatigue lifetime with logarithmic values as a preprocessing method excels when XGBoost is trained with 100% of the data. However, other normalization methods demonstrate superior accuracy in estimating test data with a 20% test and 80% train set split.

doi: *10.5829/ije.2024.37.07a.09*

Graphical Abstract



*Corresponding Author Email: m_azadi@semnan.ac.ir (M. Azadi)

Please cite this article as: Matin M, Azadi M. Effect of Training Data Ratio and Normalizing on Fatigue Lifetime Prediction of Aluminum Alloys with Machine Learning. International Journal of Engineering, Transactions A: Basics. 2024;37(07):1296-305.

1. INTRODUCTION

Fatigue loading and the resulting crack propagation are significant issues in the industry (1, 2). Researchers have attempted to increase the fatigue lifetime of industrial equipment and materials by employing various methods, including geometry-related technologies such as autofrettage (3), and manufacturing procedures such as welding (4). As a result, numerous researchers have attempted to improve the fatigue behavior of materials, geometries and estimate their fatigue lifetime. This representation is valid in the realm of alloy and metal fabrication. The following paragraph will address this claim.

First, the mechanical behavior of metals and alloys will be discussed, with a focus on aluminum alloys in the literature. Azadi et al. (5) investigated how the dwell time, thermomechanical loading factor, and maximum temperature affected the thermomechanical fatigue behavior of aluminum alloys. Furthermore, research efforts have expanded to investigate the effects of various aging heat treatments on the hardness of aluminum alloy (6). Akhtar et al. (7) investigated the optimal heat treatment temperature for aluminum alloy, revealing that 175°C was the most effective. Azadi and Parast (8) demonstrated in their work that the fatigue lifetime of aluminum alloys depends on the variation of the stress, fretting force, heat treatment condition, nanoparticles, and lubrication. Subsequently, following a thorough examination of the literature on the mechanical properties of materials, particularly aluminum alloys, it is clear that artificial intelligence (AI) methods could be helpful in estimating fatigue lifetime, providing motivation for further research.

After discussing the dependence of fatigue lifetimes on various variables and motivating using AI methods in the fatigue lifetime estimation in the second step, the literature provides the following insights.

Choi (9) estimated the steel fatigue lifetime of various chemical compositions using five different machine-learning methods. Based on his findings, XGBoost had the best accuracy with a determination coefficient (R^2) of 98.50%. Moreover, due to the substantial variability in the fatigue lifetime of alloys, particularly in the case of aluminum piston alloys, researchers are compelled to predict their fatigue lifetime behaviors and explore alternative innovative methods for estimation against destructive testing (10).

Matin and Azadi (11) conducted a Shapley Additive Explanations (SHAP) value-based analysis with XGBoost, employing five different machine learning models to predict the fatigue lifetime of aluminum alloys. Among those models, XGBoost emerged as the most effective, demonstrating compatibility with their dataset when trained on all available data.

This study motivates the researchers to find the best model for predicting the fatigue lifetime of aluminum alloys based on testing data and preprocessing methods. Furthermore, because of the potential of this work on the possibility of increasing fatigue lifetime, the literature provides the following paragraph for discussion: firstly, the normalization and scaling approaches, as preprocessing steps, transform raw data into a standardized format. For this aim, several researches have been conducted to suggest normalization methods, such as Min-Max normalization (12), Manhattan normalization (MN) (13), Euclidean normalization (EN) (14), and maximum absolute normalization (MAN) (15). Furthermore, the variability in fatigue properties arises from different sources, such as experimental conditions, material properties, and testing equipment. As a result, normalizing fatigue and its characteristics may provide valuable approaches for improving fatigue modeling (16, 17). Second, numerous studies show that the amount of training data and the ratio of training to testing data affect the accuracy of machine learning models. Additionally, these factors play a significant role in improving machine learning prediction modeling. Medar et al. (18)conducted six trials using various training ratios (ranging from 2/12 to 12/12) to calculate the mean absolute error. Ramezan et al. (19) studied the impacts of different quantities of training data on diverse supervised machine-learning classification methods.

The present study demonstrates the impact of normalization approaches on enhancing the fatigue lifetime prediction of AlSi12CuNiMg aluminum alloy, in the engine piston application. The prediction relies on stress levels in the rotary bending fatigue tests, heat treatment conditions of manufactured standard test specimens, the fretting (wear) force during the test, the corrosion time subjected to test specimens, and the existence of the lubrication during the test, all serving as inputs for the XGBoost machine learning model.

The novelty of this work lies in finding the best modeling process for estimating the fatigue lifetime in different training and testing sets through data splitting by incorporating various normalization approaches. The main application of this work is estimating the fatigue lifetime of aluminum alloys with highly accurate methods under multiple inputs, which traditional methods such as S-N curves do not provide. Moreover, by examination of test data on estimation of the fatigue lifetime, the proficiency of this method is demonstrated in reducing destructive tests like rotary bending fatigue tests for analysis.

2. RESEARCH METHODS

2. 1. Experimental Dataset This segment is based on the experimental dataset described by Azadi and Parast (8). The goal was to evaluate how different ISO

1143 standard specimens perform in corrosion fatigue, pure fatigue, and fretting fatigue tests with varying inputs in this work (8). Furthermore, these specimens were manufactured using a commercially known alloy called AlSi12CuNiMg, which has widespread application in the automotive manufacturing industry. Each sample in this dataset contains six distinct variables that serve as features (all of the variables were considered integers). The variables included "stress" ranging from 90 MPa to 120 MPa; "fretting force" with values ranging from 0 N to 20 N; the presence of "lubrication" denoted by a value of 0 for non-existing and 1 for existing; the percentage of "nano-particles" in the manufacturing of specimens, ranging from 0 to 1%; the corrosion time of manufactured specimens in H_2SO_4, ranging from 0 hours to 200 hours; and the existence of T6 heat-treatment, with a value of 0 for non-existing and 1 for existing heat-treatment. Moreover, the target is the fatigue lifetime, ranging from 500 cycles to 1,398,100 cycles.

2. 2. Modeling Techniques XGBoost is an expandable tree-boosting technique, offering significant potency and speed for machine learning tasks. Equation 1 denotes the objective involving minimizing regularization (20).

$$Obj = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{j=1}^{K} \Omega(f_k) \quad (1)$$

where $l(y_i, \hat{y}_i)$ represents the difference between the estimated and actual values, $\Omega(fj)$ and K are identified as regularization factors and the cumulative count of trees, respectively.

Various normalization and scaling methods are utilized to define a specific data range to enhance the performance and accuracy of the machine learning model. Some of these methods are as follows (21, 22):

MN: In this normalization approach, X_n is the normalized variable, X represents the unprocessed variable, and $|X|_1$represents the Manhattan norm. Equation 2 exemplifies this technique (22).

$$X_n = \frac{X}{|X|_1}$$
$$|X|_1 = \sqrt{X_1^{\,2} + X_2^{\,2} + \ldots + X_n^{\,2}} \quad (2)$$

EN: Within this normalization strategy, X_n illustrates the normalized variable, X represents the unprocessed variable, and $|X|_2$ represents the Euclidean norm. Equation 3 exemplifies this technique (22).

$$X_n = \frac{X}{|X|_2}$$
$$|X|_2 = |X_1| + |X_2| + |X_3| + \ldots + |X_n| \quad (3)$$

MAN: In this method, X_n is the normalized variable, X represents the unprocessed variable, and $MAX(|X|)$ indicates the maximum absolute value of that particular variable among all the samples. Equation 4 illustrates this method (21).

$$X_n = \frac{X}{Max(|X|)} \quad (4)$$

Modified min-max normalization (MMN): This method is closely based on Min-Max normalization. The only difference is that when calculating the logarithm value within this normalization approach, it becomes necessary to eliminate zero values from the normalized variables. Equation 5 defines this method with the following equation. Moreover, X_n represents the normalized variable, X is the unprocessed variable, X_{min} represents the minimum value among all unprocessed variables, and X_{max} denotes the maximum value among all unprocessed variables (21).

$$X_n = \frac{X - X_{\min}}{X_{Max} - X_{\min}} \text{ and } X_n \neq 0 \quad (5)$$

The performance and accuracy of estimation, regression, and machine learning modeling can be evaluated using metrics like R^2 and *RMSE*. Additionally, the *SB* offers a valuable way to illustrate a factor that covers the complete range of actual lifetimes versus the estimated fatigue lifetime (23, 24). The R^2 metric, ranging from 0 to 1, is commonly employed to assess the efficacy of a machine learning approach. A higher R^2 indicates better prediction of the target variable. Similarly, SB plots represent the actual and predicted values on the axes on a logarithmic scale. The line with the equation of $y=x$ signifies a high accuracy in modeling (experimental and predicted values), and *SB* values are the slopes of the lines that enclose the data points. A lower *SB* indicates high accuracy and less scattering of the data from the $y=x$ line (23, 24). Figure 1 provides a detailed representation of the SB plot, as depicted.

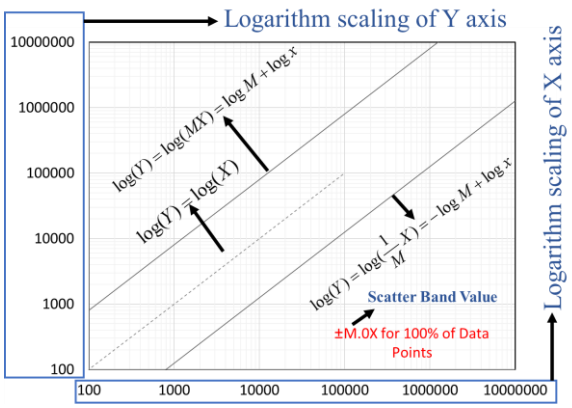


Figure 1. The detailed depiction of the SB plot

3. RESULTS AND DISCUSSION

Table 1 presents the comparative results of diverse normalization methods, employing XGBoost as the machine learning algorithm. It displays R^2 and *RMSE* values for normalized fatigue lifetime and its logarithm, using the entire dataset without splitting it into separate training and test data. Additionally, a study was conducted to predict the fatigue lifetime under various conditions using XGBoost and 100% of the data as training data (24). Their results represented the R^2 value of 97.66%, which closely aligned with the findings of the present study, where the R^2 value was 95.66%. Moreover, the results presented in this table demonstrate that MMN achieves the highest modeling accuracy among the other normalization techniques. However, it is worth noting that the fatigue lifetime without normalization was fitted remarkably well using XGBoost. Moreover, according to alternate researche, EN had a lower performance versus MMN, which, within this study, is also demonstrated to have a lower performance (25).

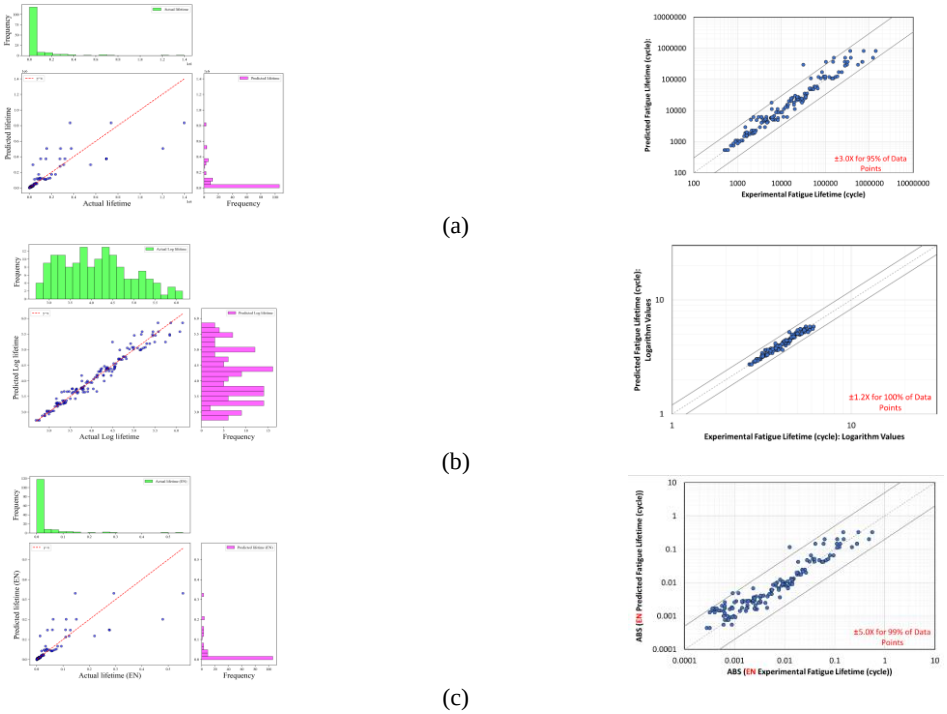
Singh and Singh (26) examined how normalization methods influenced 21 widely recognized datasets, including Iris, Australian, Breast Cancer, and others. They assessed their model's accuracy and found that, notably, in some datasets, normalization without feature selection and weighting yielded inadequate results. Notably, in some cases, this method resulted in lower accuracy. The study also revealed a fluctuating accuracy pattern across the 21 different datasets. This pattern highlighted that, in general, max absolute normalization (MAN), followed by min-max normalization, and unnormalized methods achieved higher accuracy, as indicated by their respective performances (26). Notably, the current study shows that when considering 100% of the training data, modified MMN outperforms both unnormalized and MAN methods.

Figure 2 illustrates the histograms, scatter plots, and *SB* for preprocessed fatigue lifetime data and its estimation.

TABLE 1. The values of R^2 and *RMSE* for different normalization methods using 100% of data as training data

Normalization method	Lifetime		Logarithm of lifetime	
	R^2	*RMSE*	R^2	*RMSE*
No Normalization	68.10	108624.986	95.66	0.170
MN	67.97	0.009	95.66	0.170
EN	68.09	0.043	95.66	0.170
MMN	**68.10**	**0.077**	**95.99**	**0.182**
MAN	68.10	0.077	95.66	0.170

Note:The bold values signify the superior accomplishments



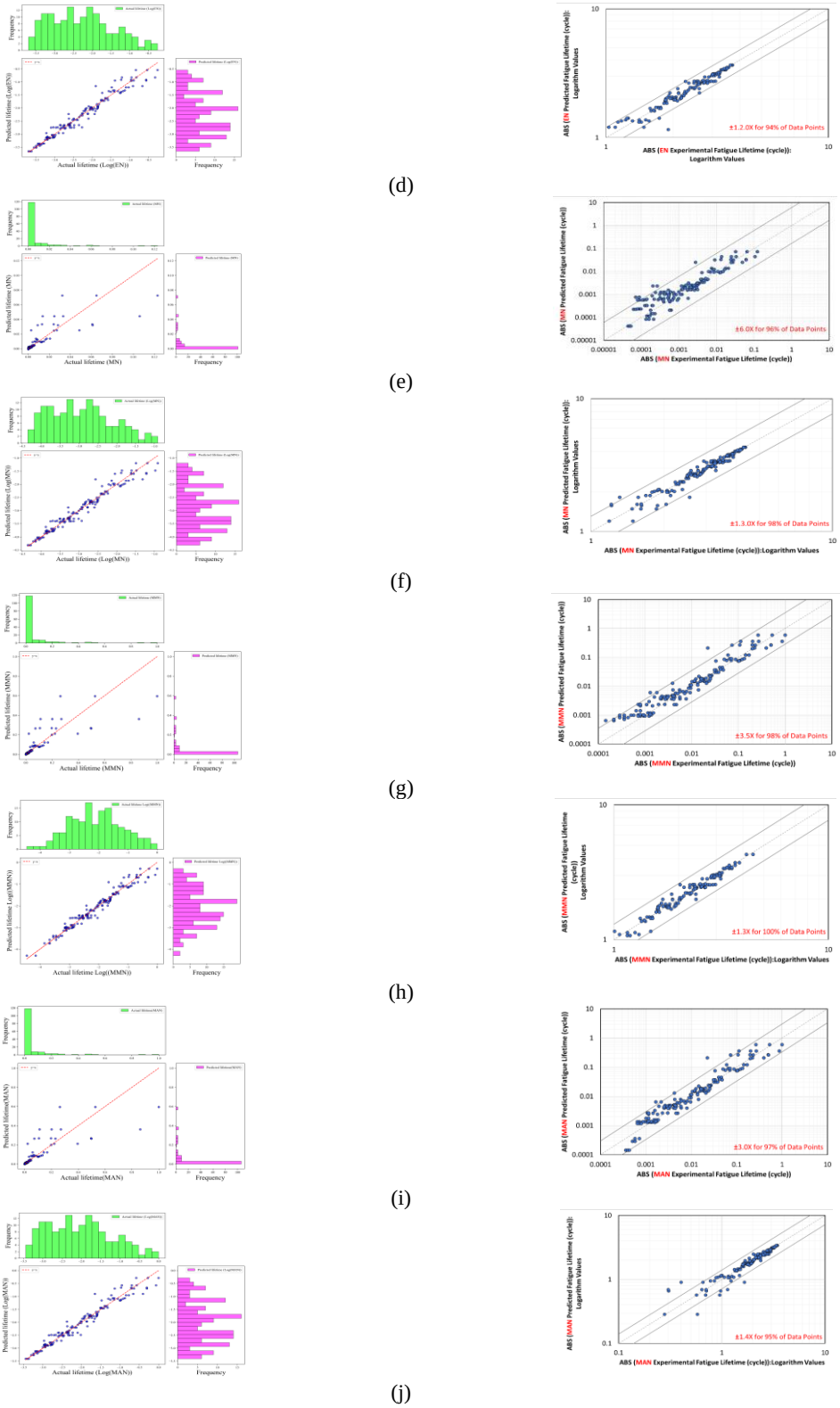


Figure 2. The histogram, scatter plot, and SB for preprocessing fatigue lifetime values and estimating them with 100% of the training data for different models: (a) No normalization of fatigue lifetime modeling, (b) logarithm value of fatigue lifetime modeling, (c) EN of fatigue lifetime modeling, (d) logarithm value of EN fatigue lifetime modeling, (e) MN of fatigue lifetime modeling, (f) logarithm value of MN fatigue lifetime modeling, (g) MMN of fatigue lifetime modeling, (h) logarithm value of MMN fatigue lifetime modeling, (i) MAN of fatigue lifetime modeling, and (j) logarithm value of MAN fatigue lifetime modeling

TABLE 2. The mean values of R^2 and *RMSE* for different normalization methods using 80-20% of the data as random training and testing

Normalization method	Lifetime		Logarithm of lifetime	
	Mean R^2	Mean *RMSE*	Mean R^2	Mean *RMSE*
No Normalization	13.95	176235.565	88.47	0.279
MN	**19.37**	**0.014**	88.67	0.276
EN	18.18	0.068	88.63	0.277
MMN	18.10	0.122	**88.72**	**0.300**
MAN	18.11	0.122	88.61	0.277

Note: The bold values signify the superior accomplishments.

Each plot on the left and right sides of the figure corresponds to a specific preprocessing method using 100% of the data for training. The figure presents the distribution of fatigue lifetime data following each preprocessing approach and its associated estimation. Moreover, a study depicts the SB plot for this experimental dataset, using 100% of the data as training data to predict the logarithmic values of fatigue lifetime, which range between 2.69897 and 6.14554 (8). This representation is included in Figure 2(b) of this work.

Several methods exist for splitting data into training and test sets. Surono et al. (27) employed a convolutional neural network to achieve an 80%-20% ratio for training and testing in various machine learning methods to estimate lung disease. In a different study, Kurdthongmee (28) investigated the impact of varying the number of training data points from 100 to 250 to determine the optimal amount for estimating parawood pith. For a more specific focus on fatigue lifetime estimation, Choi (9) reported a 30%-70% test and train ratio using XGBoost for estimating steel fatigue lifetime based on stress. The achieved R^2 for testing fatigue lifetime was 98.03%. He et al. (29) used stress and fatigue lifetime datasets in their work. They explored the effects of three different test-train ratios on estimating the fatigue lifetime of three commercial steels, concluding that a 10%-90% training ratio with artificial neural networks and random forest yielded the best results.

Table 2 in this study compares the results of various normalization methods using XGBoost as the machine learning algorithm. It calculates the average R^2 and *RMSE* over 20 iterations after randomly splitting the data into 80% training and 20% testing sets. The results demonstrate that, unlike when using 100% of the training data, when using 80% of the data for training and calculating mean R^2 for test data, all of the normalization methods improved their accuracy in modeling both the preprocessed fatigue lifetime and its logarithm.

This section of the study illustrates that when using the same algorithms and an equal number of training samples, the accuracy varies between repetitions that use different training samples, which is consistent with the other study (19). Furthermore, a comparison of the results in Tables 1 and 2 reveals that the number of training data samples has a significant impact on the metrics values. This observation is consistent with the findings of other research studies (18, 30).

Figure 3 illustrates histograms, scatter plots, and *SB* for preprocessed fatigue lifetime data and its estimation. Each plot on the left and right sides of the figure corresponds to a specific preprocessing method. These methods trained on 80% of the data and generated plots for a specific random state, which was consistent across all modeling. Moreover, the graphs relate to the test data. The figure presents the distribution of fatigue lifetime data following each preprocessing approach and its associated estimation. Furthermore, it illustrates a subset of five logarithmic normalization method approaches among all preprocessing techniques, as the remaining techniques lack satisfactory accuracy.

According to *SB* plots in Figures 2 and 3, coverages of the data by *SB* lines are different. These *SB* values in these figures correspond to optimized positions of *SB* lines for high data coverage with a low value of *SB*. Therefore, Table 3 represents *SB* values for different data coverages of *SB* lines to compare *SB* values effortlessly. Moreover, it indicates that *SB* values for fatigue lifetime and its logarithm, with no requirement for preprocessing techniques, are the best achievements.

A separate study conducted experiments using 27 different specimens subjected to varying stress levels (31). To compare the SB values presented in Table 3 for estimating the fatigue lifetime in this study with those from the other research (31). Those specimens were utilized to develop an experimental equation for the fatigue lifetime prediction, and the *SB* value was reported with a margin of ±2 for their modeling of the fatigue lifetime in cast iron crankshafts (31). In contrast, the present study used the entire dataset for training and obtained an SB value of ±10, implying that its modeling may not be as accurate as other research. Notably, the current dataset is five times larger than the previous one. Even after accounting for 90% coverage of SB lines, the SB value remained constant at ±2, demonstrating the strength and reliability of the proposed approach even when applied to over 100 data points.

The estimated fatigue lifetimes in low-cycle fatigue were not valid when the logarithmic transformer was not applied to the fatigue lifetime and its normalized values in this study. The estimated fatigue lifetimes of aluminum alloys deviated significantly from the physics of the lifetime, resulting in some negative estimated lifetimes. However, using a logarithmic transformer

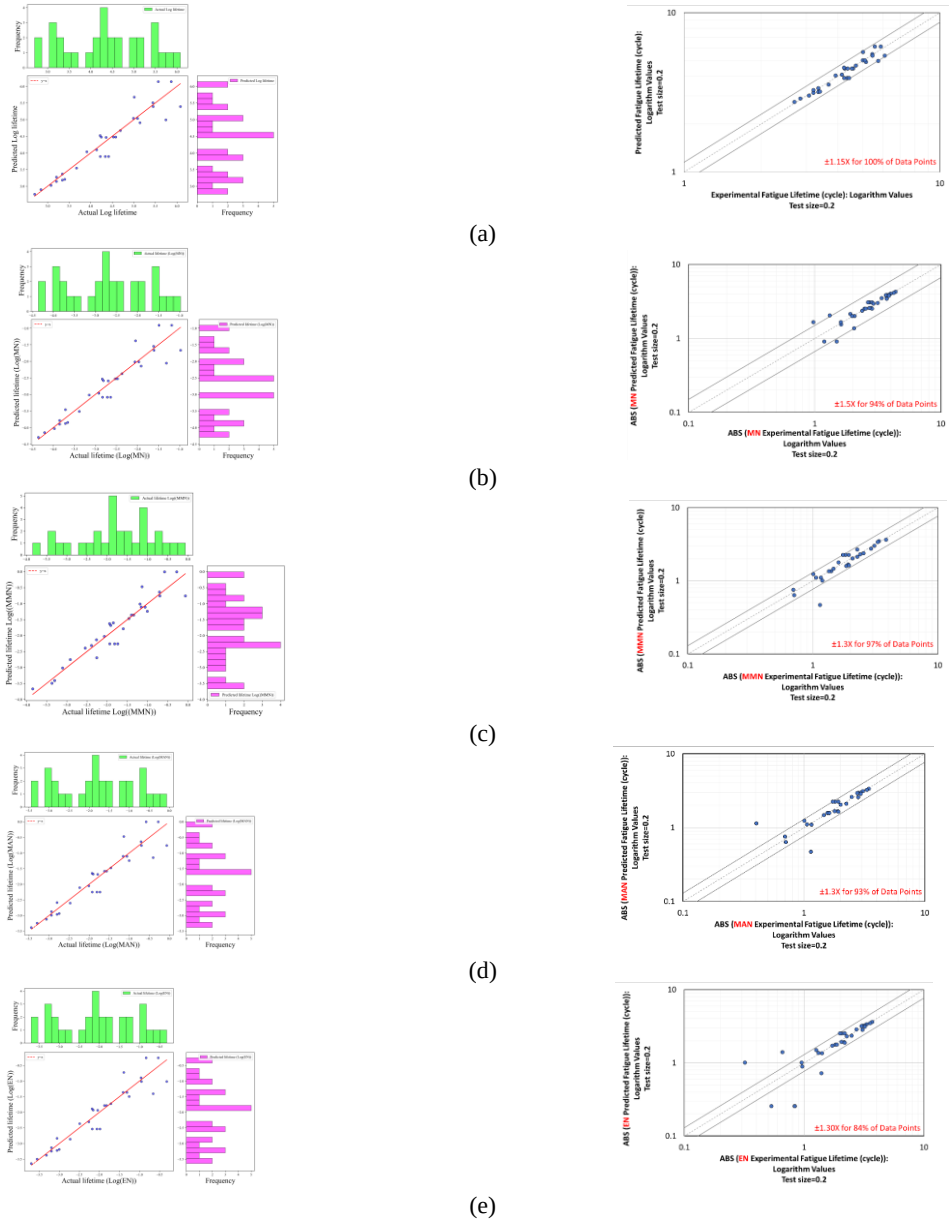


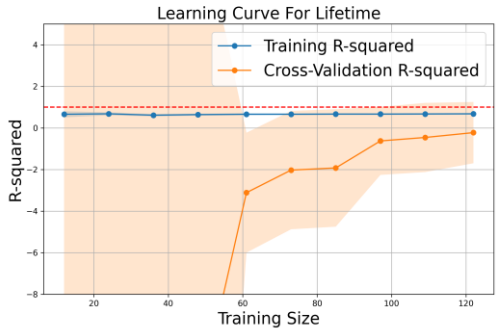
Figure 3. The histogram, scatter plot, and *SB* for same random state testing preprocessing fatigue lifetime values and estimating them with 80-20% of the data as training and testing for different models: (a) Logarithm value of fatigue lifetime modeling, (b) logarithm value of MN fatigue lifetime modeling, (c) logarithm value of MMN fatigue lifetime modeling, (d) logarithm value of MAN fatigue lifetime modeling, and (e) Logarithm value of EN fatigue lifetime modeling

proved to be a wise decision since it effectively controlled the pattern of the fatigue lifetime, transitioning from low-cycle fatigue lifetimes to high-cycle fatigue lifetimes. Furthermore, no physical laws or empirical equations related to the fatigue lifetime were used to obtain these predicted values.

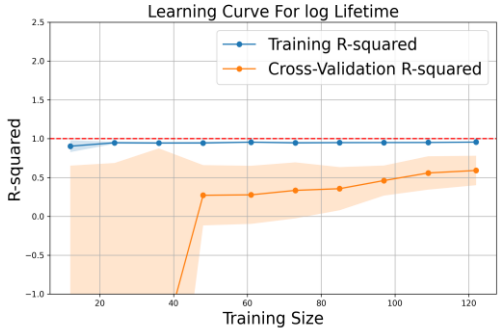
A major challenge in training or fine-tuning machine learning models is calculating the number of observations required for the optimum performance. Although having more training observations leads to better model performance, in theory, the procedure of gathering more data is typically time-consuming, expensive, or even impossible (32). Similarly, in the rotary bending fatigue tests, the preparation of specimens incurs costs for the manufacturer, and the tests themselves are both destructive and time-consuming. In such cases, learning curves can be used to determine whether the number of samples used is enough. Figure 4 represents learning curves for fatigue lifetime modeling and its logarithm value modeling. The training R^2 lines

TABLE 3. *SB* values for all methods showcasing different data coverage of *SB* lines

Method	Test size (%)	Number of scatter data	*SB*			
			100% data	95% data	90% data	85% data
Life time	0	147	10.0	3.0	2.0	1.8
Log lifetime	**0**	**147**	**1.2**	**1.1**	**1.1**	**1.1**
MAN	0	147	9.0	2.8	2.2	1.9
Log MAN	0	147	1.3	1.2	1.1	1.1
MMN	0	146	9.0	3.4	3.0	2.5
Log MMN	0	146	1.3	1.2	1.1	1.1
EN	0	147	9.0	4.3	3.5	2.9
Log EN	0	147	1.6	1.2	1.1	1.1
MN	0	147	12.0	6.0	4.0	3.5
Log MN	0	147	1.4	1.2	1.1	1.1
Log lifetime	**20**	**30**	**1.1**	**1.1**	**1.1**	**1.1**
Log MAN	20	30	2.8	1.4	1.3	1.3
Log MMN	20	30	2.4	1.3	1.2	1.2
Log EN	20	30	3.2	2.1	2.0	1.3
Log MN	20	30	1.7	1.5	1.4	1.3



(a)



(b)

Figure 4. Learning curves for (a) fatigue lifetime modeling and (b) logarithm of fatigue lifetime modeling

are roughly equal to 1, indicating that the models are catching the patterns in the training data.

Additionally, the R^2 values for cross-validation (CV) are approaching convergence with the R^2 values from the training set, as depicted in the learning curves. Generally, one of the reasons for the proximity of the training R^2 line to 1 is the overfitting of the model. However, the increasing number of training data indicates that the model requires more samples to achieve better performance and accuracy (33, 34). In this study, the accuracy of the test data, especially for logarithmic modeling, was notable and the models did not demonstrate insensitivity to small changes in the training data. Therefore, based on Figure 4, to achieve a better performance in predicting outcomes, surpassing the representations of the models presented in this study, it is recommended to expand the dataset to include additional samples.

4. CONCLUSIONS

This study focused on using various normalization methods for fatigue lifetime and its logarithm as preprocessing tasks, employing extreme gradient boosting (XGBoost) for prediction. The results included scatter bands, metrics (R^2 and *RMSE*), histograms, and scatter plots for 100% and 80% of the data employed as training. The predicted values of preprocessed fatigue lifetimes in various XGBoost models depend on specific conditions from the rotary bending fatigue tests and standard manufactured specimens used in the same test. The subsequent results were generated through this process:

- The best scatter band (*SB)* achievement for all normalization approaches, using 100% of the data for training data modeling and a random 20% of data for testing data modeling, was the logarithm of fatigue lifetime with *SB* values of ±1.2 and ±1.1, respectively. Therefore, it demonstrates the most accurate model among the other models in estimating.
- When using 100% of the data as training, all techniques of the normalization and unnormalized approaches had nearly identical values (with a difference of less than 0.5%). The SB values, on the other hand, varied. They ranged from ±9.0 to ±12.0 for non-logarithmic models and from ±1.2 to ±1.6 for logarithmic models. Therefore, it demonstrates the normalization method had no significant impact on 100% training ratio modelling.
- Calculating the mean values of R^2 for 80% of random training modeling illustrates that non-logarithmic normalization methods exhibited better accuracy (at least 4% higher than unnormalized fatigue lifetime modeling), whereas, for logarithmic

modeling, the values were closely aligned (ranging from 88.47% to 88.72%).

- Learning curves show that as the amount of training data increases, the prediction of testing data improves. As a result, additional samples should be added to the dataset to improve the accuracy of all models.

The primary limitation of the current work occurs when the logarithmic transformer is not applied to the normalized fatigue lifetime and the fatigue lifetime without normalization in XGBoost modeling. The predicted values for the fatigue lifetime and its normalization differed from the physics of the fatigue lifetime, particularly for low-cycle fatigue lifetimes, with an incorrect negative estimation. However, the logarithmic models were accurate, and the predictions agreed with the physics of the fatigue lifetime. Furthermore, for the future research aimed at improving fatigue lifetime modeling, it is suggested that an activation function be used in the output layer of a neural network. Additionally, investigating the deep learning frameworks that allow users to directly impose constraints on the layer may be beneficial, particularly for datasets with binary and categorical variables. This method is especially useful for datasets with few target values, such as the one used in this study to represent aluminum alloys. For evaluating the fatigue lifetimes of aluminum alloys in the automotive industry, these methods provide highly accurate alternatives to destructive tests such as rotary bending fatigue tests.

5. REFERENCES

1. Farrahi G, Faghidian S, Smith D. Reconstruction of residual stresses in autofrettaged thick-walled tubes from limited measurements. International Journal of Pressure Vessels and Piping. 2009;86(11):777-84. 10.1016/j.ijpvp.2009.03.010
2. Ismaiel A. Wind turbine blade dynamics simulation under the effect of atmospheric turbulence. Emerging Science Journal. 2023;7(1):162-76. 10.28991/ESJ-2023-07-01-012
3. Ali Faghidian S. Analytical approach for inverse reconstruction of eigenstrains and residual stresses in autofrettaged spherical pressure vessels. Journal of Pressure Vessel Technology. 2017;139(4):041202. 10.1115/1.4035980
4. Farrahi G, Faghidian S, Smith D. An inverse method for reconstruction of the residual stress field in welded plates. 2010. 10.1115/1.4001268
5. Azadi M, Farrahi G, Winter G, Eichlseder W. The effect of various parameters on out-of-phase thermo-mechanical fatigue lifetime of A356. 0 cast aluminum alloy. International Journal of Engineering, Transactions C: Aspects. 2013;26(12):1461-70. 10.5829/idosi.ije.2013.26.12c.06
6. Azadi M, Rezanezhad S, Zolfaghari M. Effects of various ageing heat treatments on microstructural features and hardness of piston aluminum alloy. International Journal of Engineering, Transactions A: Basics. 2019;32(1):92-8. 10.5829/ije.2019.32.01a.12
7. Akhtar M, Qamar SZ, Muhammad M, Nadeem A. Optimum heat treatment of aluminum alloy used in manufacturing of automotive piston components. Materials and Manufacturing Processes. 2018;33(16):1874-80. 10.1080/10426914.2018.1512128
8. Azadi M, Parast MSA. Data analysis of high-cycle fatigue testing on piston aluminum-silicon alloys under various conditions: Wear, lubrication, corrosion, nano-particles, heat-treating, and stress. Data in brief. 2022;41:107984. 10.1016/j.dib.2022.107984
9. Choi D-K. Data-driven materials modeling with XGBoost algorithm and statistical inference analysis for prediction of fatigue strength of steels. International Journal of Precision Engineering and Manufacturing. 2019;20:129-38. 10.1007/s12541-019-00048-6
10. Matin M, Azadi M. A Novel Machine Learning-Based Model for Predicting of Transient Fatigue Lifetime in Piston Aluminum Alloys. Available at SSRN 4598611. 10.2139/ssrn.4598611
11. Matin M, Azadi M. Machine learning-based modeling for estimating bending fatigue lifetimes in AlSi12CuNiMg aluminum alloy of engine pistons under different inputs: Fretting force, corrosion time, lubrication, heat-treating, nano-particles, and stress. Corrosion Time, Lubrication, Heat-Treating, Nano-Particles, and Stress. 2023. 10.2139/ssrn.4549376
12. Munkhdalai L, Munkhdalai T, Park KH, Lee HG, Li M, Ryu KH. Mixture of activation functions with extended min-max normalization for forex market prediction. IEEE Access. 2019;7:183680-91. 10.1109/ACCESS.2019.2959789
13. Zhou W, Liu A, Wu L, Chen X. A L1 normalization enhanced dynamic window method for SSVEP-based BCIs. Journal of Neuroscience Methods. 2022;380:109688. 10.1016/j.jneumeth.2022.109688
14. Dai Z, Chen W, Huang X, Li B, Zhu L, He L, et al., editors. Cnn descriptor improvement based on l2-normalization and feature pooling for patch classification. 2018 IEEE International Conference on Robotics and Biomimetics (ROBIO); 2018: IEEE. 10.1109/ROBIO.2018.8665330
15. Gómez-Escalonilla V, Martínez-Santos P, Martín-Loeches M. Preprocessing approaches in machine-learning-based groundwater potential mapping: an application to the Koulikoro and Bamako regions, Mali. Hydrology and Earth System Sciences. 2022;26(2):221-43. 10.5194/hess-26-221-2022
16. Lv S, Liu C, Chen D, Zheng J, You Z, You L. Normalization of fatigue characteristics for asphalt mixtures under different stress states. Construction and Building Materials. 2018;177:33-42. 10.1016/j.conbuildmat.2018.05.109
17. Lv S, Wang P, Fan X, Cabrera MB, Hu L, Peng X, et al. Normalized comparative study on fatigue characteristics of different pavement materials. Construction and Building Materials. 2021;271:121907. 10.1016/j.conbuildmat.2020.121907
18. Medar R, Rajpurohit VS, Rashmi B, editors. Impact of training and testing data splits on accuracy of time series forecasting in machine learning. 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA); 2017: IEEE. 10.1109/ICCUBEA.2017.8463779
19. Ramezan CA, Warner TA, Maxwell AE, Price BS. Effects of training set size on supervised machine-learning land-cover classification of large-area high-resolution remotely sensed data. Remote Sensing. 2021;13(3):368. 10.3390/rs13030368
20. Meng Y, Yang N, Qian Z, Zhang G. What makes an online review more helpful: an interpretation framework using XGBoost and SHAP values. Journal of Theoretical and Applied Electronic Commerce Research. 2020;16(3):466-90. 10.3390/jtaer16030029
21. Md AQ, Kulkarni S, Joshua CJ, Vaichole T, Mohan S, Iwendi C. Enhanced preprocessing approach using ensemble machine learning algorithms for detecting liver disease. Biomedicines. 2023;11(2):581. 10.3390/biomedicines11020581

22. Chiu W-Y, Chen B-S. Mobile location estimation in urban areas using mixed Manhattan/Euclidean norm and convex optimization. IEEE transactions on Wireless Communications. 2009;8(1):414-23. 10.1109/T-WC.2009.080156
23. Azadi M, Shahsavand A, Parast MSA. Analyzing experimental data from reciprocating wear testing on piston aluminum alloys, with and without clay nano-particle reinforcement. Data in Brief. 2022;45:108766. 10.1016/j.dib.2022.108766
24. Nasiri H, Azadi M, Dadashi A. Interpretable extreme gradient boosting machine learning model for fatigue lifetimes in 3D-printed polylactic acid biomaterials. Available at SSRN 4364418. 2023. 10.2139/ssrn.4364418
25. Eesa AS, Arabo WK. A normalization methods for backpropagation: a comparative study. Science Journal of University of Zakho. 2017;5(4):319-23. 10.25271/2017.5.4.381
26. Singh D, Singh B. Investigating the impact of data normalization on classification performance. Applied Soft Computing. 2020;97:105524. 10.1016/j.asoc.2019.105524
27. Surono S, Afitian MYF, Setyawan A, Arofah DKE, Thobirin A. Comparison of CNN Classification Model using Machine Learning with Bayesian Optimizer. HighTech and Innovation Journal. 2023;4(3):531-42. 10.28991/HIJ-2023-04-03-05
28. Kurdthongmee W. Comprehensive Evaluation of Deep Neural Network Architectures for Parawood Pith Estimation. HighTech and Innovation Journal. 2023;4(3):543-59. 10.28991/HIJ-2023-04-03-06
29. He L, Wang Z, Akebono H, Sugeta A. Machine learning-based predictions of fatigue life and fatigue limit for steels. Journal of Materials Science & Technology. 2021;90:9-19. 10.1016/j.jmst.2021.02.021
30. Althnian A, AlSaeed D, Al-Baity H, Samha A, Dris AB, Alzakari N, et al. Impact of dataset size on classification performance: an empirical evaluation in the medical domain. Applied Sciences. 2021;11(2):796. 10.3390/app11020796
31. Khameneh MJ, Azadi M. Evaluation of high-cycle bending fatigue and fracture behaviors in EN-GJS700-2 ductile cast iron of crankshafts. Engineering Failure Analysis. 2018;85:189-200. 10.1016/j.engfailanal.2017.12.017
32. Cruz F, Castelli M. Learning Curves Prediction for a Transformers-Based Model. Available at SSRN 4305463. 2023. 10.28991/ESJ-2023-07-05-03
33. Zhang G, Shi Y, Yin P, Liu F, Fang Y, Li X, et al. A machine learning model based on ultrasound image features to assess the risk of sentinel lymph node metastasis in breast cancer patients: Applications of scikit-learn and SHAP. Frontiers in Oncology. 2022;12:944569. 10.3389/fonc.2022.944569
34. Giola C, Danti P, Magnani S. Learning curves: A novel approach for robustness improvement of load forecasting. Engineering Proceedings. 2021;5(1):38. 10.3390/engproc2021005038

Persian Abstract

چکیده

ارزیابی برآورد طول عمر خستگی برای آلیاژهای آلومینیوم، به ویژه در صنعت خودرو، امری حیاتی است. این مقاله تاثیر روش‌های عادی سازی مختلف را بر عملکرد تخمین طول عمر خستگی با استفاده از تقویت گرادیان شدید (XGBoost) به عنوان یک روش یادگیری ماشین نظارت شده بررسی می‌کند. به این منظور در این مطالعه مجموعه داده ای شامل ورودی های فیزیکی و تجربی متنوع مربوط به آلیاژ آلومینیوم در کنار عمر خسته به عنوان خروجی استفاده شده است. علاوه بر این، قبل از برازش مدل XGBoost، روش‌های مختلف پیش پردازش طول عمر خستگی اعمال می‌شوند و سپس با استفاده از معیارهایی مانند خطای میانگین مربعات ریشه (RMSE)، ضریب تعیین (R^2)، و باند پراکندگی (SB) ارزیابی شده اند. نتایج به‌دست‌آمده نشان می‌دهد که مدل‌سازی طول عمر خستگی با مقادیر لگاریتمی به عنوان یک روش پیش‌پردازش زمانی که XGBoost با ۱۰۰ درصد داده‌ها آموزش داده می‌شود، برتری دارد. با این حال، سایر روش‌های نرمال‌سازی دقت بالاتری را در تخمین داده‌های آزمایش با تقسیم ۲۰٪ تست و ۸۰٪ مجموعه داده یادیگیری نشان می‌دهند.