Materials Today Communications 39 (2024) 109054

Contents lists available at ScienceDirect

Materials Today Communications

journal homepage: www.elsevier.com/locate/mtcomm

Machine learning for fatigue lifetime predictions in 3D-printed polylactic acid biomaterials based on interpretable extreme gradient boosting model

Hamid Nasiri [a], Ali Dadashi [b], Mohammad Azadi [b,*]

[a] Semnan University, Semnan, Iran
[b] Faculty of Mechanical Engineering, Semnan University, Semnan, Iran

ARTICLE INFO

Keywords:
Biomaterials
Machine learning
3D printing
Fatigue lifetime
Extreme gradient boosting

ABSTRACT

Modeling of the fatigue lifetimes in 3D-printed biomaterials of Polylactic acid (PLA) is presented in this article based on machine learning (ML) techniques of interpretable extreme gradient boosting (XGBoost) and Shapley additive explanations. For this objective, standard testing samples were additive-manufactured from PLA under different 3D printing parameters. Then, the fatigue experiments were performed on specimens under various stress levels. Based on these data, three ML methods were utilized for modeling the PLA fatigue lifetimes, including XGBoost, random forest, and support vector regression, besides a common nonlinear regression analysis. The obtained results indicated that XGBoost had superior modeling results, compared to other ML techniques and the regression analysis. The coefficient of determination was 97.66 % with a scatter-band of ±1.3, which was a narrow scatter in fatigue modeling.

1. Introduction

Additive manufacturing (AM) by 3D printing provides the possibility of producing parts with complex geometry or functional materials with layered construction [1]. Moreover, the technology causes less environmental damage due to the reduction of waste materials [2]. Therefore, AM has revolutionized the design and manufacturing processes [3]. However, there are challenges, such as porosity due to poor bonding between materials [4], deformation due to residual stresses resulting from rapid cooling [5], and anisotropy [6]. It is necessary to overcome these challenges to ensure the appropriate performance of additively manufactured parts.

Research works have been conducted to cover various aspects of printing parameters. Abeykoon et al. [7] performed a comparative study to optimize fused deposition modeling (FDM) parameters for improved Polylactic acid (PLA) and Acrylonitrile butadiene styrene (ABS) 3D-printed structures. Such investigations could be expensive and time-consuming [3]. In addition, it is extremely difficult to investigate the whole AM parameters, quickly and accurately. Data-driven models can predict the behavior of parts made with AM by processing the data collected by machine learning (ML) algorithms. These models are beneficial in decision-making [8]. ML techniques improve the efficiency of the 3D printing process by covering various aspects, including part design, manufacturing parameters, and product quality [9]. Therefore, they have been used to understand the relationship between the properties, structure, and process parameters. Kumar et al. [10] developed a ML algorithm to characterize various stages of the 3D printing process by Long Short-Term Memory Algorithm (LSTM), suitable for PLA, ABS, and PETG (polyethylene terephthalate glycol) filaments. The model enabled the decision support in improving energy and time inefficiencies in 3D printing. Moreover, different studies have utilized machine learning techniques in FDM 3D printing of the PLA material for the purpose of minimizing the support waste [11] and predicting the geometric accuracy [12].

In the area of additive manufacturing, optimizing the performance of printed materials is an ongoing challenge, and understanding the relationship between printing conditions and material behavior is of paramount importance.

The fatigue behavior of PLA components fabricated with 3D-printing has been studied in the existing literature. Decreasing the nozzle diameter increased the fatigue lifetime, enhancing the structural density and resolution [13]. Conversely, larger diameters promoted the interlayer cohesion but reduced the surface quality [14]. The effect of printing speed was reported to vary with the nozzle diameter. However, higher print speeds improved the tensile properties by enhancing bonding and reducing the print time [15]. The increased nozzle

* Corresponding author.
E-mail address: m_azadi@semnan.ac.ir (M. Azadi).

https://doi.org/10.1016/j.mtcomm.2024.109054
Received 12 February 2024; Received in revised form 17 April 2024; Accepted 27 April 2024
Available online 28 April 2024

temperature reduced the fatigue properties: for instance, samples printed at 240°C exhibited the lowest lifetime due to the material overflow near the melting point [13]. The optimal parameters for ABS-enhanced PLA were 215°C, 90 mm/s of the speed, and 100 % of filling [7]. Notably, the tensile strength of 3D-printed PEEK varied with the nozzle temperature, while PEEK (polyether ether ketone) outperformed PEI (polyethylenimine) in this regard [16].

However, there is a lack of models considering fabrication parameters on fatigue properties of the material [17]. Most of the existing models for fatigue of PLA are based on empirical or simplified approaches that have several limitations neglecting the effects of process parameters. These models may not be able to capture the complex and nonlinear fatigue mechanisms and failure modes of additively manufactured products, especially under long-term service conditions. Moreover, the existing machine learning models for fatigue of PLA also have some drawbacks, such as overfitting, lack of interpretability, and insufficient use of runout data. Researches have been conducted to address these challenges [18–20]. In this study, a new machine learning model that can accurately and reliably predict the fatigue properties of PLA, based on the print parameters using advanced experimental techniques. Such a model would enable design and optimization of PLA products with the enhanced fatigue performance and durability [21]. The extreme gradient boosting (XGBoost) algorithm employed in this research incorporates regularization techniques to mitigate the risk of overfitting.

Sai et al. [22] demonstrated the promising performance of popular ML techniques, including Random Forest (RF), Support Vector Machine (SVM), and boosting algorithms like GBoost and XGBoost in estimating the fatigue lives within a reasonable margin of error. They also explored the application of machine learning algorithms for predicting the fatigue phenomenon in the glass fiber reinforced aluminum laminate [23]. This study aimed to solve the complex interaction between 3D printing parameters and the fatigue lifetime of additively-manufactured PLA biomaterial samples. The fatigue lifetime, a crucial mechanical property, signifies the durability and reliability of these samples when subjected to cyclic loading conditions. By using the power of ML models, a comprehensive exploration to estimate the fatigue lifetime with different 3D printing conditions can be provided. The primary goal in this study was to explore which printing parameters significantly influence the fatigue lifetime, providing valuable insights into strategies for enhancing the fatigue resistance of 3D-printed PLA specimens. Through modeling, analysis, and comparison, factors that influenced PLA fatigue lifetimes were introduced, bridging the gap between 3D printing technology and material performance prediction.

The novelty of this work lies in its pioneering attempt to find out the behavior of 3D-printed biomaterials, specifically PLA, using advanced ML techniques. By coupling the power of interpretable extreme gradient boosting (XGBoost) and Shapley additive explanations, this study takes a remarkable leap forward in understanding the intricate world of PLA fatigue lifetimes. Such insights help to understand the way performance and durability of 3D-printed biomaterials can be optimized.

2. Materials and methods

2.1. Experimental dataset

The data for this study was collected through a systematic process designed to investigate the impact of various 3D printing parameters on the bending fatigue properties of standard PLA specimens. The goal was to gain a comprehensive understanding of the mechanical behavior of these specimens under different printing conditions. The total number of 162 standard dog-bone samples were fabricated in accordance with the ISO-1143 standard as shown in Fig. 1. The standard is designed for metallic materials but it is also used in this study due to the absence of a standard for fatigue testing method of additive manufactured polymers. A specific setup was designed for FDM process, ensuring consistent and

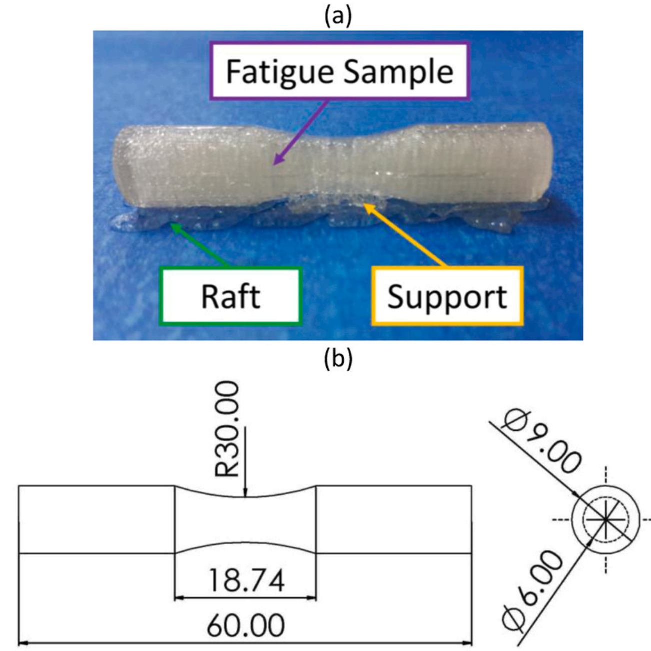

Fig. 1. The (a) printed and (b) schematic geometry of dog-bone standard samples.

accurate printing of samples [13]. The samples were 3D printed using the FDM method and subjected to cyclic bending loads through fatigue tests. The transparent PLA filaments with a diameter of 1.75 mm, sourced from YouSu Company, China, were utilized for 3D printing. The choice of PLA as the base material was made to represent a commonly used polymer in FDM processes. [24,25].

To examine the influence of specific 3D printing parameters on the bending fatigue properties of the PLA specimens, a controlled experimental setup was employed. This fatigue testing was done under the load-controlled fully-reversed cyclic condition, using a rotary bending fatigue testing device, manufactured by Santam Company, Iran. According to Fig. 2, the device can apply bending stresses to the rotating specimen at a 100 Hz of frequency. The specimens were tested in the high-cycle fatigue (HCF) regime at room temperature and under 2.5–17.5 MPa of the bending stress [24,25]. These stress levels refer to the magnitude of cyclic stress applied to the specimens.

The FDM printer was employed for the fabrication of test specimens. The printer capabilities allowed us to precisely control key printing parameters, namely the print speed, nozzle diameter, and nozzle temperature. These variables were selected to span a range of values in order to assess their impact on fatigue behavior [12,13].

- Print speed: varied from 5 to 15 mm/s.
- Nozzle diameter: ranged from 0.2 to 0.6 mm.
- Nozzle temperature: set within the range of 180–240°C.

The data collection process involved systematically printing and testing the specimens under different combinations of the mentioned printing parameters. Each combination was replicated to ensure statistical significance. For each unique set of printing parameters, a series of dog-bone specimens were 3D-printed. These specimens were then subjected to bending fatigue tests to evaluate their mechanical properties. Table 1 lists the statistical description of the dataset. Notably, 162 specimens were fabricated and tested; however, 47 outliers were excluded from the analysis due to high scatter bands (or variations). Therefore, the dataset consists of 115 samples and four features. The preliminary analysis suggests that the out of scatter-band data might be attributed to defects arising from the poor quality in the 3D-printed

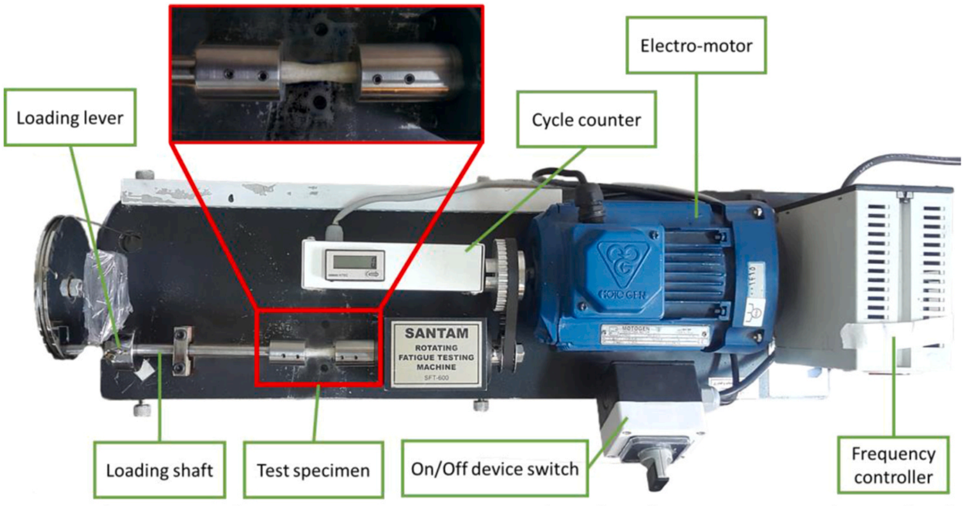

Fig. 2. The image of SFT-600 Santam rotating bending fatigue testing device [16].

Table 1
The statistical description of the dataset.

Variable	Min	Max	Mean	STD
Nozzle Diameter	0.20	0.60	0.41	0.16
Print Speed	5.00	15.00	10.17	4.13
Nozzle Temperature	180.00	240.00	209.48	24.16
Stress Level	2.50	17.50	8.59	3.52
Average Fatigue Lifetime	1400.00	1,500,000.00	128,857.00	338,018.70

specimens [24]. Table 1 summarizes the distribution of the data and provides a concise representation of its central tendency and spread. As suggested in the literature [26], a dataset should have at least 10 samples for every variable fed into a regression model. More details of such regression analysis could be found in the literature [24,25].

2.2. Modeling techniques

Model interpretability is a critical issue for ML techniques. Lundberg and Lee [27] suggested the Shapley additive explanation (SHAP) to solve this problem. SHAP uses Shapley values to interpret the predictions of ML models [26]. The Shapley values measure the importance of each feature in predicting an outcome. Shapley values (ϕ_i) can be calculated, as follows,

$$\phi_i = \sum\nolimits_{S \epsilon N \backslash \{i\}} \frac{|S|!(M-|S|-1)!}{M!}[f(S \cup \{i\}) - f(S)] \tag{1}$$

where f represents the model, M is the number of variables, N denotes the set of all inputs, and S is a subset of M that excludes the i-th variable from N [28]. Shapley values offer insights into the importance and impact of each feature on the model output. By considering all possible combinations of features and their contributions, SHAP values estimate the fair allocation of the prediction outcome among the features. They calculate the marginal contribution of each feature to the prediction compared to all other possible combinations of features. A positive SHAP value for a feature indicates that the presence of that feature increases the prediction value, while a negative value suggests its absence decreases the prediction. The magnitude of the SHAP value represents the extent of influence a feature has on the prediction. Features with larger absolute SHAP values have a more significant impact on the outcome [26].

XGBoost is a highly efficient, scalable implementation of gradient-boosting decision trees (GBDT). The loss function employed by XGBoost uses second-order Taylor expansion, compared to first-order one in GBDT [29]. This issue allows XGBoost to produce more accurate results than GBDT [30,31]. Moreover, XGBoost uses regularization, which helps reduce the risk of overfitting. Regularization adds a penalty to the objective, which reduces the complexity of the model, preventing it from memorizing or fitting too closely to the training data [32]. The objective is shown in Eq. (2) [33],

$$\phi = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{j=1}^{K} \Omega(f_j) \tag{2}$$

where $l(y_i, \hat{y}_i)$ denotes the difference between the output ($\hat{y}_i$) and the target value (y_i). Moreover, $\Omega(f_j)$ and K represent the penalty term and the number of regression trees, respectively [33].

Random forest (RF) is a nonparametric ensemble learning technique capable of solving regression and classification problems [34]. This approach generates multiple decision trees on random subsets of the data, allowing each tree to use a different feature subset. The RF model then aggregates the predictions of each tree, resulting in an improved generalization performance [35,36].

RF builds an ensemble of N decision trees. Then, the trees calculate the final output for an input sample ($\mathbf{x}$), based on Eq. (3) [37],

$$\hat{T}(\mathbf{x}) = \frac{1}{N}\sum_{n=1}^{N} \hat{T}_n(\mathbf{x}) \tag{3}$$

where $\hat{T}_n(\mathbf{x})$ represents the n-th tree output [37].

Support vector regression (SVR) is a supervised ML technique. SVR is based on the idea of creating a regression model that not only fits the data well but also maximizes the generalization performance by finding the most appropriate hyperplane to separate the data samples [38].

To compare the ML methods with usual regression modeling, the ANOVA analysis was used with Design-Expert [39,40]. The confidence level was 95 % in this non-linear regression modeling with a natural logarithmic function and two-way interactions. Using a sensitivity analysis, F-Value and P-Value were reported plus the coefficient of determination (R^2). Another examiner of modeling was the scatter-band (SB), which shows a factor that cover the whole data of experimental lifetimes versus prediction values. The width of the SB indicates the level of agreement between the experiments and predictions; a narrower band suggests a more accurate model, while a wider band indicates greater disparity.

3. Results and discussion

Based on Fig. 3, the experimental fatigue lifetimes of the manufactured PLA specimens under different 3D-printing parameters are shown, such as the layer thickness, infill density, and print speed. Additional information regarding the dataset and parameters was published, providing more detailed insights [24]. The samples, out of scatter band points, which significantly deviated from the general trend, were filtered in this S-N curve. In addition, the run-out tests, which are the tests that did not result in failure even after 1.5 million cycles, were indicated by an arrow next to the corresponding point. These tests were terminated since they exceeded the failure criterion.

Fig. 4 depicts the Pearson correlation between variables for linear and nonlinear considerations. In Fig. 4(a), the strongest linear correlations involve the fatigue lifetime, with the nozzle diameter and stress level showing moderate negative linear correlations. Moreover, the stress level had a positive correlation with the print speed and a negative correlation with the nozzle temperature. This issue could imply that these parameters were significant predictors for the fatigue lifetime. The relationships between other parameters were generally weak to moderate, suggesting that while there was some linear association, it was not particularly strong. Incorporating nonlinear considerations into the analysis, as shown in Fig. 4(b), revealed that not only the individual parameters were important, but also their squared and interaction terms played a significant role in the behavior of the system.

Table 2 summarizes the parameters used for each model. The first model, XGBoost, employs a base learner in the form of gradient boosted trees with an exact greedy tree construction algorithm. In total, 100 gradient boosted trees are used in this model. The second model, RF, comprises 20 trees and requires a minimum of 2 samples to split an internal node. The quality of splits in the RF model is evaluated using the Squared Error function. SVR, as the third model, adopts a radial basis function kernel and sets the tolerance for the stopping criterion at 0.001, with an epsilon value of 0.1. Moreover, Table 3 presents the comparison results of various techniques, including 20 independent runs and the value of averages and standard deviations. The root mean squared error (RMSE), R^2, and SB were reported. The dataset was randomly split into a training set and a testing set using an 80:20 ratio, with 80 % of the data allocated for training and 20 % for testing. For a fair comparison between different models, the same train and test sets were used for all techniques. Optimizing hyperparameters is essential for refining machine learning models, as they dictate the model behavior and performance. A methodical tuning process can enhance accuracy and the ability to generalize. Furthermore, while the SVR model uses a radial basis function kernel, investigating alternative kernels is crucial. Different kernels can detect unique data patterns, potentially improving model efficacy. Such exploration might uncover a kernel that optimally balances bias and variance, boosting predictive strength.

The obtained results implied that XGBoost could accurately estimate the average fatigue lifetime based on various printing parameters, which outperformed other methods. Moreover, to find the superiority of models, a two-tailed Welch's statistical t-test was considered for RMSE between XGBoost and other techniques, with a confidence level of 95 % [41]. Welch's t-test is a nonparametric univariate statistical test, useful when the two samples have unequal variances [42]. As Table 3 illustrates, the null hypothesis was rejected based on P-Value (lower than 0.05) and therefore, the results were statistically meaningful. Scatter band (SB) is a statistical tool used to assess the reliability of predictions. As shown in Fig. 5, by comparing the predicted values with the actual fatigue lifetime, the parameter quantifies the dispersion, providing insights into the accuracy of models and their predictive performance. For XGBoost, a ± 1.3 range indicated predictions were closely aligned with actuals, suggesting high model accuracy. The RF model also showed a ± 1.3 variance, however the lower R^2 value implied less precision. Nonlinear regression and SVR with tight bands of ± 1.4 scatter band, had lower performance.

The findings presented in the mentioned table, suggest that the number of experimental specimens was indeed sufficient. The successful performance of these models in analyzing the data and generating meaningful results underscored the robustness of the dataset and its suitability for the research objectives. The ML techniques, XGBoost and RF, along with the nonlinear regression modeling, produced highly precise and statistically significant results, as indicated by their high R^2 values. The outcome demonstrates the robustness and reliability of the dataset, indicating that the number of specimens was logical and provided sufficient information to draw meaningful and statistically

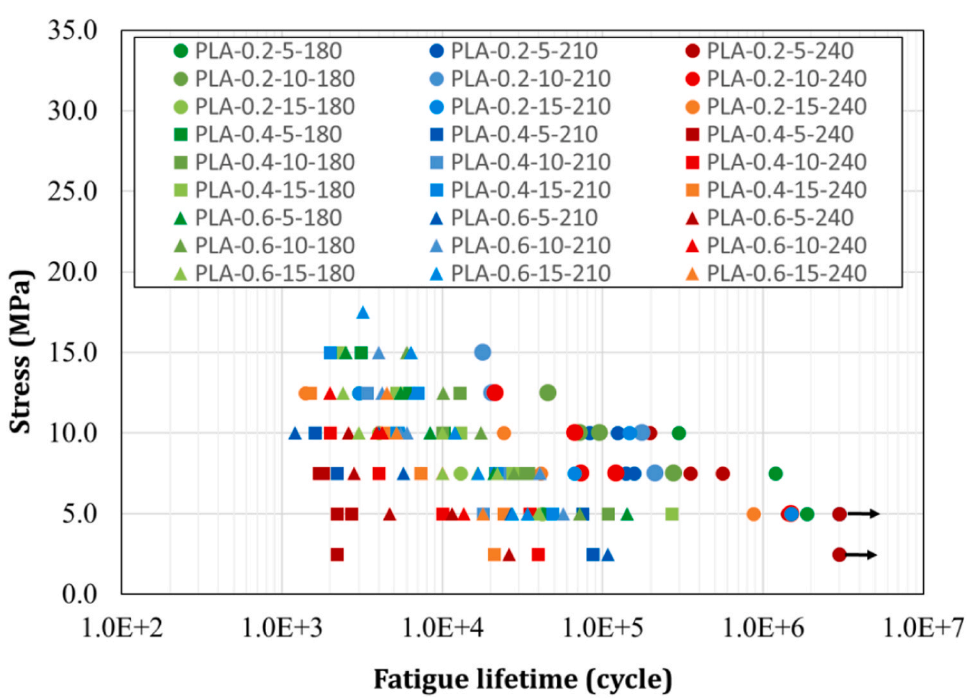

Fig. 3. The S-N diagrams for PLA samples, 3D-printed with different process parameters.

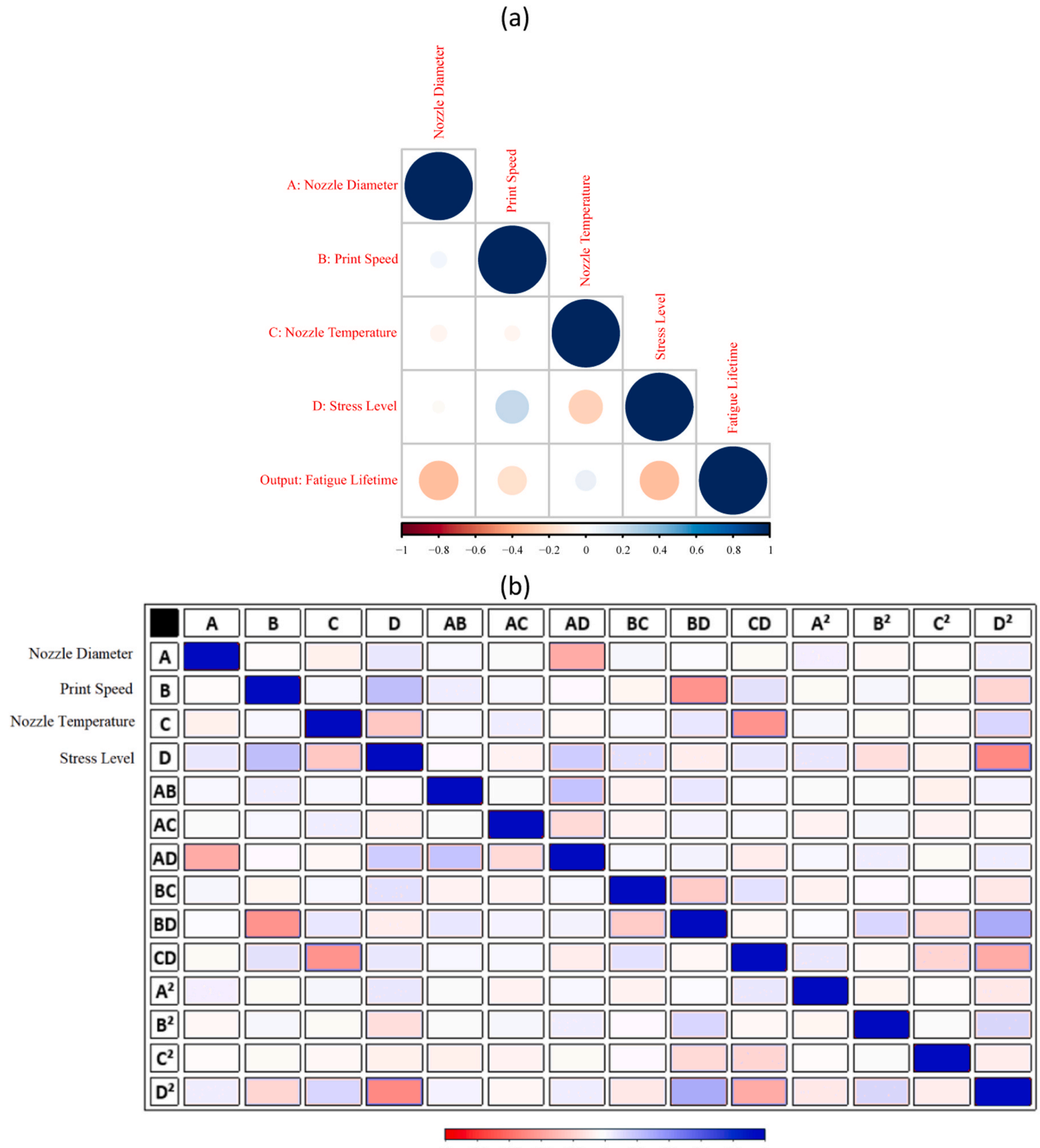

Fig. 4. The Pearson correlation between variables for (a) linear and (b) nonlinear considerations.

significant conclusions. The small standard deviations further supported the consistency and accuracy of the results across multiple runs.

To comprehend the impact of a single variable on the model output, the SHAP value of that variable versus the parameter value for all samples is demonstrated (Fig. 6). The x-axis represents the value of each parameter, and the y-axis represents the SHAP value, which measures how much each parameter contributes to fatigue lifetime. The color of the samples indicates the interaction effect of another parameter that has the strongest influence on the SHAP value. For example, the red dots in Fig. 6(a) shows that there was no correlation between the nozzle temperature and stress level. The plots reveal the complex and nonlinear relationships between the printing parameters and the stress level, as well as the potential interactions among them. Moreover, based on Fig. 6 (b), there is no correlation between the nozzle diameter and the stress level.

Based on the differences between actual (experimental) and predicted values of the average fatigue lifetimes of PLA by ML methods, XGBoost had the lowest average error. Then, SHAP was applied to XGBoost model (i.e., the best model based on RMSE and R^2). As an agreement, this issue was also reported for XGBoost in the literature [28], which can significantly boost both the estimation performance and also the computational speed.

Zhou et al. [32] indicated that XGBoost, light gradient boosting machine (LightGBM), and categorical boosting (CatBoost) outperformed the RF model in predicting outcomes when using the synthetic minority oversampling technique (SMOTE) oversampling. Specifically, the prediction results of the retention fraction were significantly lower than the other models, with the R^2 value of 0.8530 and the mean square error (MSE) of 0.0165. On the other hand, XGBoost, LightGBM, and CatBoost achieved good regression prediction effects, with R^2 greater than 0.93 and MSE less than 0.008. These findings suggested that XGBoost, LightGBM and CatBoost were more suitable for predicting outcomes,

Table 2
Model parameters used in ML techniques for predictions.

Parameters	Values
Model: XGBoost	
Base Learner	Gradient boosted tree
Tree construction algorithm	Exact greedy
Learning rate	0.3
Lagrange multiplier	0
Number of gradient boosted trees	100
Model: Random Forest (RF)	
Number of trees	20
Minimum number of samples required to split an internal node	2
The function to measure the quality of a split	Squared Error
Model: Support Vector Regression (SVR)	
Kernel	Radial basis function
Tolerance for stopping criterion	0.001
Epsilon	0.1

Table 3
Comparison results of ML techniques and the regression analysis.

Modeling techniques	RMSE	R^2	P-value	SB
ML: XGBoost	**47,937.18 (15,332.28)**	**0.9766 (0.0117)**	**-**	**±1.3**
ML: RF	75,244.05 (22,988.28)	0.9430 (0.0307)	1.00E-4	±1.3
ML: SVR	245,254.53 (57,251.50)	0.4488 (0.0773)	7.13E-13	±1.4
Nonlinear regression modeling	-	0.8120	<0.0001	±1.4

Note: The bolded text is the best model. Moreover, the values in parenthesis are the standard deviations for 20 times of running

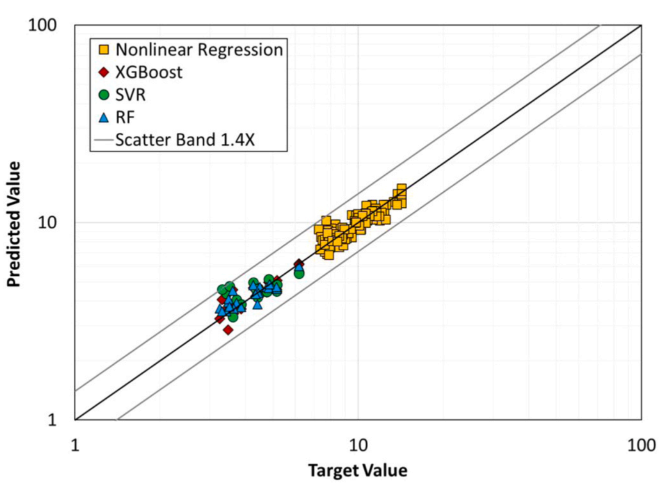

Fig. 5. The scatter-band analysis of different models.

when using SMOTE oversampling than the RF model. The differences between XGBoost, gradient boosting, RF, and LightGBM were statistically insignificant [44,45]. What sets XGBoost apart is its versatility when it comes to hyper-parameter tuning, particularly its randomization hyper-parameters like subsampling rate and the number of features selected at each split. By setting these values appropriately, Bentejac et al. [46] demonstrated that the XGBoost performance can be significantly improved while reducing the hyper-parameter search complexity, making it a compelling choice for those seeking a powerful and adaptable gradient boosting algorithm.

Based on the literature [36], it was found that both XGBoost-based and RF-based models had a superior predictive performance, compared to baseline models. The order of prediction accuracy was XGBoost, RF, support vector machine (SVM), multi-layer perceptron (MLP), and multivariate adaptive regression spline (MARS). However, when comparing the performance of XGBoost and RF, XGBoost was found to be superior to RF in terms of predictive accuracy. Notably, the XGBoost algorithm is based on the decision trees and gradient boosting, which allows it to handle complex data and achieve a better accuracy than other models.

Additionally, XGBoost was shown to be superior to RF and SVR models in predicting outcomes in complex scenarios [33,47]. Specifically, XGBoost was able to accurately estimate the results based on monitored operational parameters with significantly higher accuracy than the other two models. This issue suggested that XGBoost was better suited for predicting outcomes in complex scenarios, which were often encountered in various industries. Moreover, XGBoost was shown to significantly improve both computational speed and performance [43]. This note was due in part to its advanced algorithms and optimization techniques, which enabled faster training and inference times, compared to other machine learning models. XGBoost is also capable of handling missing data and feature interactions, which can increase the model accuracy. Generally, the superior performance of XGBoost may be attributed to its advanced boosting algorithms, which enable efficient model training on complex datasets. Additionally, the ability of XGBoost to handle missing data and feature interactions contributes to its superior performance. These findings highlighted the potential of XGBoost as a reliable and effective tool for predicting outcomes in various fields where accuracy was of utmost importance. However, it is essential to recognize that the performance of ML algorithms appeared to be case-dependent. Further research is required to explore the full potential of XGBoost and its applications across different domains, taking into account the unique characteristics and requirements of each case.

Variables were ranked by SHAP values according to their level of importance for predicting the average fatigue lifetime (Fig. 7(a) and (b)). Results demonstrated that the nozzle diameter and the print speed had the highest and lowest effectiveness for predicting the fatigue lifetime, respectively. Considering mean SHAP values indicated that except for the nozzle temperature, which had a positive impact on the fatigue lifetime, other parameters showed a negative effect on the output. There was considerable agreement between Pearson correlation assessments (Fig. 4) and SHAP (Fig. 7). As it can be seen in Fig. 4(a), the nozzle diameter had the highest correlation with the fatigue lifetime. In addition, in Fig. 7(a), the nozzle diameter had the highest mean SHAP value.

This negative correlation between the nozzle temperature and fatigue lifetime can be attributed to the layer overflow induced by higher temperatures specifically near the melting point of material [13]. Based on the results in Fig. 7, the print speed has the lowest mean SHAP value, which had agreement with the regression analysis, where the P-Value of this factor was obtained as 0.7775 (higher than 0.05, non-significant factor). Other variables were significant with P-Value less than 0.05. Besides, the importance order of these parameters was the stress level (F-Value: 110.68), the nozzle diameter (F-Value: 87.48), and the nozzle temperature (F-Value: 18.79). This result was different from the SHAP analysis for the first-ranked important parameter, which was the nozzle diameter.

One possible cause of the difference between the model analyses is the assumption of the feature independence that SHAP values make. On the other hand, the regression coefficients do not make such an assumption and directly measure the effect of a unit change in a feature on the prediction, holding all other features constant. In other words, SHAP values may not fully account for the interdependencies between features, which can lead to an overestimation or underestimation of a feature influence, whereas regression coefficients consider the change in context with other variables held fixed. Consequently, regression coefficients and SHAP values may yield different importance rankings when features are correlated. One way to address the differences between the regression and SHAP analyses in design is to use both methods as complementary tools for understanding the effects of the input

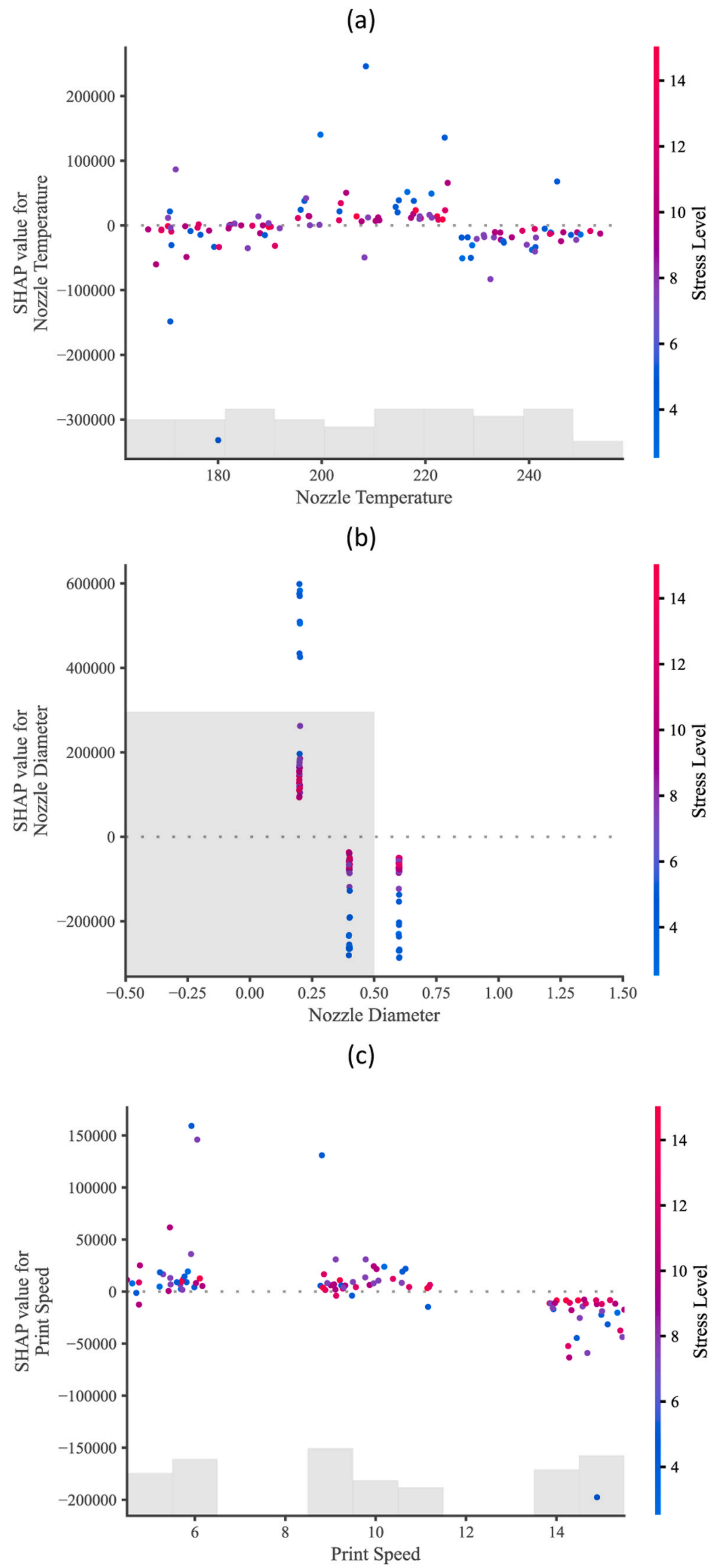

Fig. 6. SHAP scatter plots for different printing parameters besides the stress level.

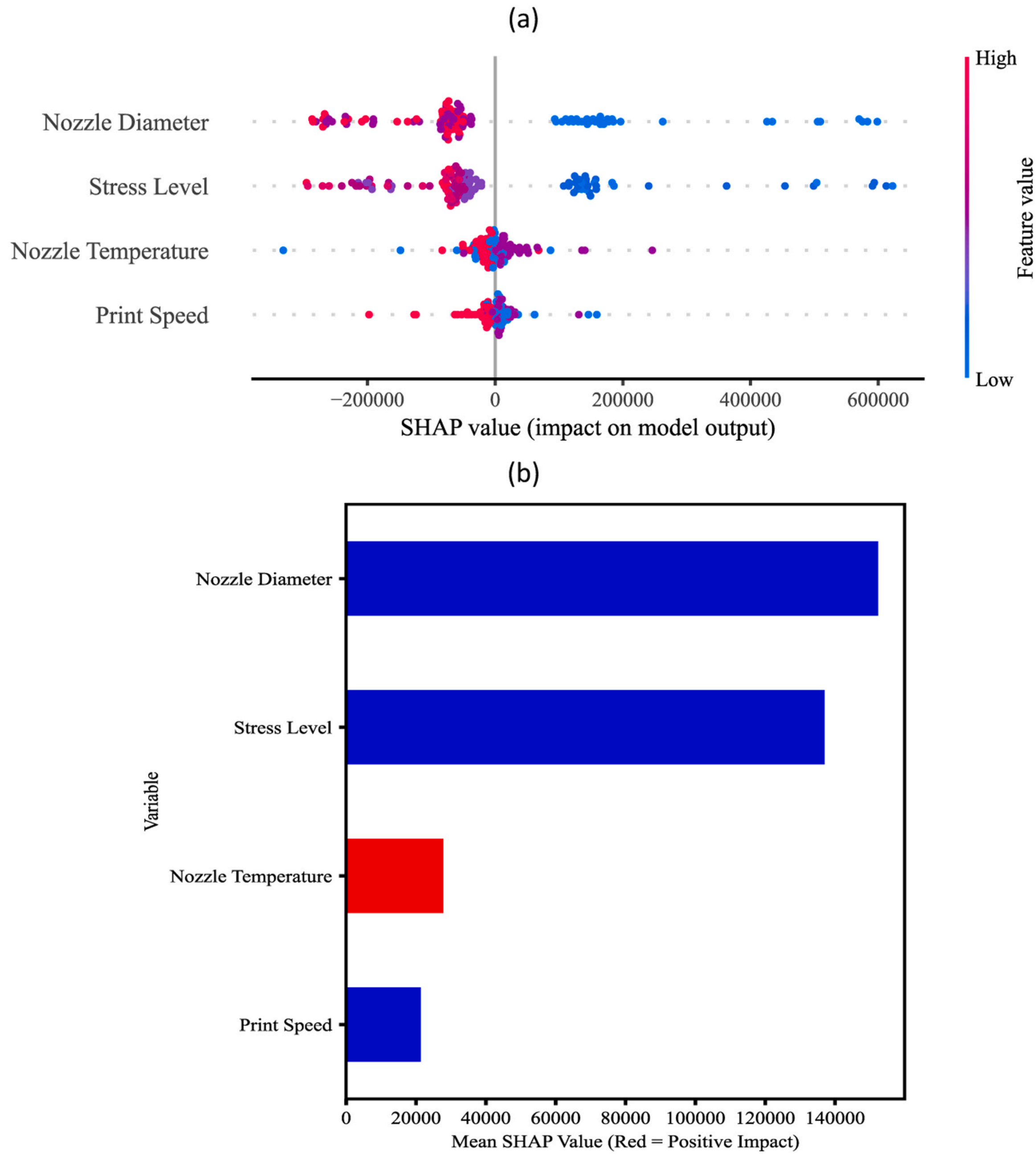

Fig. 7. ML prediction results: (a) SHAP beeswarm and (b) bar plots.

parameters on the PLA fatigue lifetimes. By comparing the outcomes of the insights from SHAP and regression analyses, one can identify whether the parameters have significant or insignificant influence on PLA fatigue lifetimes. This comparison can guide model selection, inform feature engineering, and optimize design processes.

4. Conclusions

In this study, machine learning models were used to estimate the fatigue lifetime of additive-manufactured PLA samples under various 3D printing conditions. The results of this study provided valuable insights into the relationship between 3D printing conditions and the fatigue lifetime of PLA samples. The following results were found from modeling can be summarized,

- The superior model for predicting the fatigue lifetime of PLA samples was found to be XGBoost, with RMSE of 47,937.18, R^2 of 97.66 %, and the scatter-band of ±1.3. These metrics indicated that the XGBoost model provided accurate and reliable predictions of PLA fatigue lifetimes.
- The SHAP analysis was conducted to determine the relative importance of various 3D printing conditions on the fatigue lifetime of PLA specimens. This analysis revealed that the nozzle diameter had the highest influence on the fatigue lifetime, while the 3D printing speed had the lowest effect. This issue suggested that optimizing the nozzle diameter could be an effective strategy for improving the fatigue resistance of 3D-printed PLA samples.
- A usual regression analysis was also performed to compare the results with the SHAP analysis. The result showed that the print speed was a non-significant factor in predicting PLA fatigue lifetimes,

indicating that it had a little impact on the fatigue resistance of PLA specimens.

- The order of important parameters in estimating PLA fatigue lifetimes differed between the regression and SHAP analyses. In the regression analysis, the stress level was the most important factor, followed by the nozzle diameter and the nozzle temperature. However, in the SHAP analysis, the nozzle diameter was found to be the most important factor, followed by the stress level and the nozzle temperature. This behavior suggested that different factors may be more or less important depending on the specific ML models used and the analysis technique employed.

CRediT authorship contribution statement

Hamid Nasiri: Writing – original draft, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ali Dadashi:** Writing – original draft, Visualization, Software, Resources, Methodology, Investigation, Formal analysis, Data curation. **Mohammad Azadi:** Writing – review & editing, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data will be made available on request.

References

[1] W. Schneller, M. Leitner, S. Pomberger, F. Grün, S. Leuders, T. Pfeifer, O. Jantschner, Fatigue strength assessment of additively manufactured metallic structures considering bulk and surface layer characteristics, Addit. Manuf. Vol. 40 (2021) 101930.
[2] M. Javaid, A. Haleem, R.P. Singh, R. Suman, S. Rab, Role of additive manufacturing applications towards environmental sustainability, Adv. Ind. Eng. Polym. Res. Vol. 4 (4) (2021) 312–322.
[3] O. Diegel, A. Nordin, D. Motte, A Practical Guide to Design for Additive Manufacturing, Springer Singapore, 2019.
[4] A. Dadash, M. Azadi, Multi-objective numerical optimization of 3D-printed polylactic acid bio-metamaterial based on topology, filling pattern, and infill density via fatigue lifetime and mass, PLOS ONE Vol. 18 (9) (2023) e0291021.
[5] D. Jia, F. Li, Y. Zhang, 3D-printing process design of lattice compressor impeller based on residual stress and deformation, *2020 10:1*, Sci. Rep. Vol. 10 (1) (2020) 1–11.
[6] N.S. Hmeidat, R.C. Pack, S.J. Talley, R.B. Moore, B.G. Compton, Mechanical anisotropy in polymer composites produced by material extrusion additive manufacturing, Addit. Manuf. Vol. 34 (2020) 101385.
[7] C. Abeykoon, P. Sri-Amphorn, A. Fernando, Optimization of fused deposition modeling parameters for improved PLA and ABS 3D printed structures, Int. J. Lightweight Mater. Manuf. Vol. 3 (3) (2020) 284–297.
[8] S. Kumar, T. Gopi, N. Harikeerthana, M.K. Gupta, V. Gaur, G.M. Krolczyk, C.S. Wu, Machine learning techniques in additive manufacturing: a state of the art review on design, processes and production control, *2022 34:1*, J. Intell. Manuf. Vol. 34 (1) (2022) 21–55.
[9] Z. Jin, Z. Zhang, K. Demir, G.X. Gu, Machine learning for advanced additive manufacturing, Matter Vol. 3 (5) (2020) 1541–1556.
[10] R. Kumar, R. Ghosh, R. Malik, K.S. Sangwan, C. Herrmann, Development of machine learning algorithm for characterization and estimation of energy consumption of various stages during 3D printing, Procedia CIRP Vol. 107 (2022) 65–70.
[11] J. Jiang, G. Hu, X. Li, X. Xu, P. Zheng, J. Stringer, Analysis and prediction of printable bridge length in fused deposition modelling based on back propagation neural network, https://doi.org/10.1080/17452759.2019.1576010, Vol. 14, No. 3, pp. 253–266, 2019.
[12] N. Decker, M. Lyu, Y. Wang, Q. Huang, Geometric accuracy prediction and improvement for additive manufacturing using triangular mesh shape data, J. Manuf. Sci. Eng. Trans. ASME Vol. 143 (6) (2021).
[13] A. Dadashi, M. Azadi, Optimization of 3D printing parameters in polylactic acid bio-metamaterial under cyclic bending loading considering fracture features, Heliyon Vol. 10 (4) (2024) e26357.
[14] P. Wang, B. Zou, H. Xiao, S. Ding, C. Huang, Effects of printing parameters of fused deposition modeling on mechanical properties, surface quality, and microstructure of PEEK, J. Mater. Process. Technol. Vol. 271 (2019) 62–74.
[15] A.A. Ansari, M. Kamil, Effect of print speed and extrusion temperature on properties of 3D printed PLA using fused deposition modeling process, Mater. Today.: Proc. Vol. 45 (2021) 5462–5468.
[16] S. Ding, B. Zou, P. Wang, H. Ding, Effects of nozzle temperature and building orientation on mechanical properties and microstructure of PEEK and PEI printed by 3D-FDM, Polym. Test. Vol. 78 (2019) 105948.
[17] G. Morettini, M. Palmieri, L. Capponi, L. Landi, Comprehensive characterization of mechanical and physical properties of PLA structures printed by FFF-3D-printing process in different directions, Prog. Addit. Manuf. Vol. 7 (2022) 1111–1122.
[18] S. Alkunte, I. Fidan, Machine learning-based fatigue life prediction of functionally graded materials using material extrusion technology, J. Compos. Sci. Vol. 7 (10) (2023) 420.
[19] S. Gao, X. Yue, H. Wang, Predictability of different machine learning approaches on the fatigue life of additive-manufactured porous titanium structure, Metals Vol. 14 (3) (2024) 320.
[20] G. Liu, C. Dobbins, M. D'Souza, N. Phuong, A machine learning approach for detecting fatigue during repetitive physical tasks, Pers. Ubiquitous Comput. Vol. 27 (2023) 2103–2120.
[21] J. Chen, Y. Liu, Probabilistic physics-guided machine learning for fatigue data analysis, Expert Syst. Appl. Vol. 168 (2021) 114316.
[22] N.J. Sai, P. Rathore, A. Chauhan, Machine learning-based predictions of fatigue life for multi-principal element alloys, Scr. Mater. Vol. 226 (2023) 115214.
[23] W. Sai, G.B. Chai, N. Srikanth, Fatigue life prediction of GLARE composites using regression tree ensemble-based machine learning model, Adv. Theory Simul. Vol. 3 (No. 6) (2020) 2000048.
[24] M. Azadi, A. Dadashi, Experimental fatigue dataset for additive-manufactured 3D-printed Polylactic acid biomaterials under fully-reversed rotating-bending bending loadings, Data Brief. Vol. 41 (2022) 107846.
[25] A. Dadashi, M. Azadi, Experimental bending fatigue data of additive-manufactured PLA biomaterial fabricated by different 3D printing parameters, Prog. Addit. Manuf. (2022).
[26] R.D. Riley, K.I.E. Snell, J. Ensor, D.L. Burke, F.E. Harrell, K.G.M. Moons, G. S. Collins, Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes, Stat. Med. Vol. 38 (No. 7) (2019) 1276–1296.
[27] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst. Vol. 30 (2017) 4765–4774.
[28] X. Wen, Y. Xie, L. Wu, L. Jiang, Quantifying and comparing the effects of key risk factors on various types of roadway segment crashes with LightGBM and SHAP, Accid. Anal. \ Prev. Vol. 159 (2021) 106261.
[29] M.R. Abbasniya, S.A. Sheikholeslamzadeh, H. Nasiri, S. Emami, Classification of breast tumors based on histopathology images using deep features and ensemble of gradient boosting methods, Comput. Electr. Eng. Vol. 103 (2022) 108382.
[30] K. Song, F. Yan, T. Ding, L. Gao, S. Lu, A steel property optimization model based on the XGBoost algorithm and improved PSO, Comput. Mater. Sci. Vol. 174 (2020) 109472.
[31] H. Nasiri, S. Hasani, Automated detection of COVID-19 cases from chest X-ray images using deep neural network and XGBoost, Radiography Vol. 28 (3) (2022) 732–738.
[32] K. Zhou, S. Li, X. Zhou, Y. Hu, C. Zhang, J. Liu, Data-driven prediction and analysis method for nanoparticle transport behavior in porous media, Measurement Vol. 172 (2021) 108869.
[33] A. Maleki, M. Raahemi, H. Nasiri, Breast cancer diagnosis from histopathology images using deep neural network and XGBoost, Biomed. Signal Process. Control Vol. 86 (2023) 105152.
[34] A. Farzipour, R. Elmi, H. Nasiri, Detection of Monkeypox cases based on symptoms using XGBoost and shapley additive explanations methods, Diagnostics Vol. 13 (14) (2023) 2391.
[35] H. Gong, Y. Sun, X. Shu, B. Huang, Use of random forests regression for predicting IRI of asphalt pavements, Constr. Build. Mater. Vol. 189 (2018) 890–897.
[36] W. Zhang, C. Wu, H. Zhong, Y. Li, L. Wang, Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization, Geosci. Front. Vol. 12 (1) (2021) 469–477.
[37] R. Fatahi, H. Nasiri, E. Dadfar, S. Chehreh Chelgani, Modeling of energy consumption factors for an industrial cement vertical roller mill by SHAP-XGBoost: a "conscious lab" approach, Sci. Rep. Vol. 12 (1) (2022) 7543.
[38] M. Premalatha, C.V. Lakshmi, Svm trade-off between maximize the margin and minimize the variables used for regression, Int. J. Pure Appl. Math. Vol. 87 (6) (2013) 741–750.
[39] M. Azadi, A. Dadashi, M.S. Aghareb Parast, S. Dezianian, A. Bagheri, A. Kami, M. Kianifar, Vahid Asghari, A comparative study for high-cycle bending fatigue lifetime and fracture behavior of extruded and additive-manufactured 3D-printed acrylonitrile butadiene styrene polymers, Int. J. Addit. Manuf. Struct. Vol. 1 (1) (2022) 1–14.
[40] M.S. Aghareb Parast, A. Bagheri, A. Kami, M. Azadi, V. Asghari, Bending fatigue behavior of fused filament fabrication 3D-printed ABS and PLA joints with rotary friction welding, Prog. Addit. Manuf. (2022) 1–17.
[41] H. Nasiri, M.M. Ebadzadeh, Multi-step-ahead stock price prediction using recurrent fuzzy neural network and variational mode decomposition, Appl. Soft Comput. Vol. 148 (2023) 110867.
[42] H. Nasiri, M.M. Ebadzadeh, MFRFNN: multi-functional recurrent fuzzy neural network for chaotic time series prediction, Neurocomputing Vol. 507 (2022) 292–310.

[43] C. Wang, X.P. Tan, S.B. Tor, C.S. Lim, Machine learning in additive manufacturing: state-of-the-art and perspectives, Addit. Manuf. Vol. 36 (2020) 101538.
[44] R.P. Sheridan, A. Liaw, M. Tudor, Light Gradient Boosting Machine as a Regression Method for Quantitative Structure-Activity Relationships, *arXivLabs*, 2021.
[45] M. Niang, M. Ndong, I. Dioum, I. Diop, M. Mashaly, M.A.A. El GhanyComparison of Random Forest and Extreme Gradient Boosting Fingerprints to Enhance an indoor Wifi Localization System, International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), No. Cairo, Egypt2021.
[46] C. Bentéjac, A. Csörgő, G. Martínez-Muñoz, A comparative analysis of gradient boosting algorithms, Artif. Intell. Rev. Vol. 54 (3) (2021) 1937–1967.
[47] P. Devan, N. Khare, An efficient XGBoost–DNN-based classification model for network intrusion detection system, Neural Comput. Appl. Vol. 32 (16) (2020) 12499–12514.