MDPI

Article

Predictability of Different Machine Learning Approaches on the Fatigue Life of Additive-Manufactured Porous Titanium Structure

Shuailong Gao, Xuezheng Yue and Hao Wang *

Interdisciplinary Centre for Additive Manufacturing (ICAM), School of Materials and Chemistry,
University of Shanghai for Science and Technology, Shanghai 200093, China; 212203143@st.usst.edu.cn (S.G.);
usst-yzyz@usst.edu.cn (X.Y.)
* Correspondence: haowang7@usst.edu.cn

Abstract: Due to their outstanding mechanical properties and biocompatibility, additively manufactured titanium porous structures are extensively utilized in the domain of medical metal implants. Implants frequently undergo cyclic loading, underscoring the significance of predicting their fatigue performance. Nevertheless, a fatigue life model tailored to additively manufactured titanium porous structures is currently absent. This study employs multiple linear regression, artificial neural networks, support vector machines, and random forests machine learning models to assess the impact of structural and mechanical factors on fatigue life. Four standard maximum likelihood models were trained, and their predictions were compared with fatigue experiments to validate the efficacy of the machine learning models. The findings suggest that the fatigue life is governed by both the fatigue stress and the overall yield stress of the porous structures. Furthermore, it is recommended that the optimal combination of hyperparameters involves setting the first hidden layer of the artificial neural network model to three or four neurons, establishing the gamma value of the support vector machine model at 0.0001 with C set to 30, and configuring the n_estimators of the random forest model to three with max_depth set to seven.

Keywords: machine learning; porous structure; titanium; additive manufacturing; fatigue

1. Introduction

Due to its high strength and low modulus, titanium is widely used in the field of medical metal implants; however, "stress shielding" causes premature implant failure [1,2]. Additive manufacturing (AM) creates porous structures with a modulus compatible with bones and enhanced tissue growth capability [3,4]. However, additively manufactured porous structures differ from traditionally manufactured counterparts on defects, microstructures, and geometry structures [5,6], which bring difficulty on property evaluation. Fortunately, the rise of machine learning (ML) in recent years has enabled the prediction of fatigue performance.

In previous studies, in order to optimize tissue regeneration, it is imperative to understand the fatigue behavior of porous implants placed in the body for long periods of time [7]. Based on different loading methods (tension–tension, tension–compression, compression–compression), the fatigue life of porous structures has been determined by Lietaert et al. [8]. According to Wycisk et al. [9], the surface roughness of porous additively fabricated structures has the greatest effect on fatigue life because they have a greater surface-to-volume ratio than blocks. It has been reported by Hrabe et al. [10] that the fatigue strength of electron beam melting (EBM) Ti-6Al-4V porous structures is significantly reduced by stress concentrations near rough surfaces. The strain rate sensitivity of porous structures is reduced by high roughness and a lower relative density [11]. In the same stress amplitude, Lindemann et al. [12–14] found that fatigue life decreased with increasing stress

ratios. According to Hooreweder et al. [15], chemical corrosion could smooth the surface of the strut, reducing the stress concentration near unit nodes and fatigue crack initiation. In previous studies, the multiple factors influencing AM porous implants' fatigue performance and the fatigue damage mechanism were not clearly understood, which makes predicting the fatigue life of porous implants difficult.

In the present day, a growing field of ML has shown great potential in determining materials' mechanical and fatigue parameters [16–18]. The most popular ML models among them are multiple linear regression (MLR), artificial neural networks (ANNs), support vector machines (SVMs), and random forests (RFs). A rising number of studies have successfully used the ML models to forecast the fatigue life of components made using AM [19–24]. In 2020, Zhan et al. [22] predicted the fatigue life of 300M-AerMet100 in additive manufacturing through a combination of experimental methods, numerical simulations, and an artificial neural network model. In 2021, Zhan et al. [20] established a data-driven analysis platform grounded on continuous damage mechanics, employing artificial neural networks, support vector machines, and random forest models for predicting the fatigue life of additively manufactured SS316L components. In 2021, Zhan et al. [21] employed random forest and artificial neural network models to forecast the fatigue life of additively manufactured Ti6Al4V, SS316L, and AlSi10Mg. In 2021, Bao et al. [24] employed support vector machine and artificial neural network models to investigate the fatigue life concerning defect location, size, and morphology in additively manufactured Ti-6Al-4V alloy. In 2023, Shi et al. [23] introduced a methodology to address the issue of data sparsity in the fatigue model and applied it to forecast the fatigue life of additively manufactured AlSi10Mg alloy. In comparison to conventional statistical approaches, they have higher computing accuracy and efficiency for nonlinear regression analysis and small sample prediction [25,26].

In the field of materials science, both machine learning and additive manufacturing are emerging directions, especially with the recent development of using machine learning models to predict the performance of additive manufacturing components. While we aim to focus on the materials themselves, particularly on key issues such as fatigue-life prediction, the limited research time and inadequate preparation of relevant datasets make it necessary to initially discuss the accuracy of different machine learning models. This work also provides an important foundation and perspective for future related research. In contrast to previous research findings, this study marks a breakthrough as it pioneers the utilization of a data-driven machine learning model to explore the fatigue life of additively manufactured porous structures. Initially, the study predicted the fatigue life through the application of four ML models: MLR, ANN, SVM, and RF. This prediction is based on a limited amount of experimental data that specifically emphasizes yield stress and fatigue stress. Then, the prediction data and models were analyzed to validate their applicability by identifying any errors or discrepancies. Finally, parametric studies were conducted to explore the impact of key parameters in the ML models on prediction performance. Additionally, the study recommends the optimal hyperparameters for each ML model type.

2. Methodology

2.1. Experimental Data

In this work, we have employed experimental data from compression fatigue tests of porous Ti2448 (Ti-24Nb-4Zr-8Sn) rhombic dodecahedron structures fabricated by electron beam powder bed fusion (Figures 6–8 in Ref. [1]). Details on the materials and tests are also available in Ref. [1].

2.2. Machine Learning Models

2.2.1. Multiple Linear Regression (MLR)

If the function curve is a straight line, then it is known as a linear regression. If the linear regression curve has more than one independent variable, then it is known as MLR. For modeling linear relationships between independent and dependent variables, MLR

is an important predictive analytics method. By fitting an optimal linear function, we can describe the relationship between multiple independent variables and the dependent variable. The diagram of the MLR regression structure is shown in Figure 1.
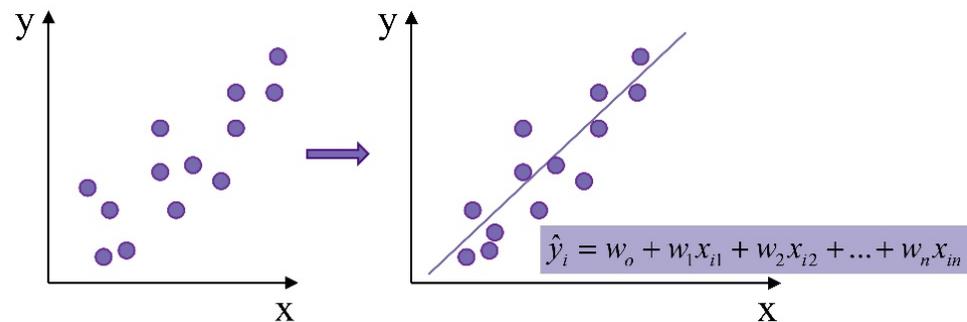
Figure 1. The diagram of the multiple linear regression model.

For sample *i* with *n* features, its regression result can be written as an equation:

$$\hat{y}_i = w_o + w_1 x_{i1} + w_2 x_{i2} + \ldots + w_n x_{in} \tag{1}$$

where w is collectively referred to as the parameters of the model, w_0 is called the intercept, and w_1 to w_n are called the regression coefficients.

In this work, the brief steps of MLR model training are shown below:

(1) In multiple linear regression, our loss function is defined as follows:

$$\sum_{i=1}^{m} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{m} (y_i - X_i w)^2 \tag{2}$$

where y_i is the true label corresponding to sample *i* and $\hat{y}_i$ is the predicted label corresponding to sample *i*. This loss function represents the squared result of the L2 paradigm for vector $y - \hat{y}$, which is essentially the Euclidean distance, i.e., it is the sum of the squares of the corresponding subtracted squares for each point on the two vectors, and then squared.

(2) Considering that the goal of multiple linear regression is to have as small of a difference between the predicted and true values as possible, the solution objective can be rephrased as follows:

$$\min_w \|y - Xw\|_2^{\,2} \tag{3}$$

2.2.2. Artificial Neural Networks (ANN)

The ANN model is a simulation of the human brain's neural network, which establishes the relationship between inputs and outputs based on how many neurons communicate and transfer information among samples. It is widely used to predict S–N relationships [27] and to map complex nonlinear relationships [28]. This is mainly because many activation functions (such as the linear, sigmoid, and hyperbolic tangent) can be used to train the ANN model [29]. The diagram of the ANN regression structure is shown in Figure 2.

In general, a typical ANN contains an input layer, a hidden layer, and an output layer, and each layer contains many neurons. During the training process of an ANN, neurons in the current layer receive signals from neurons in the previous layer and then output the signals to neurons in the next layer. The training process of a single neuron is shown below:

$$y_i = f_i(\sum w_{ij} x_j - t_i) \tag{4}$$

where y_i represents the output, x_j is the input, w_{ij} is the weight, and f_i is the activation function. When the sum $\sum w_{ij} x_j$ is larger than a limit, the output is activated by the activation function. Furthermore, the values of weight and limit could be revised for the sake of minimizing the output errors.
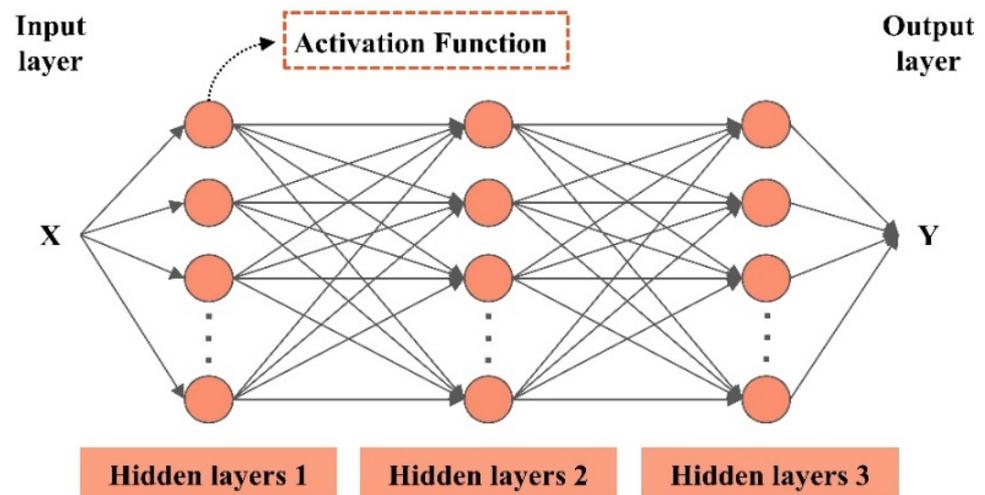
Figure 2. The diagram of the artificial neural networks model.

In this work, the commonly used back propagation algorithm and gradient descent optimization algorithm are used to train the ANN and the brief steps are shown below:

(1) Initialize all the weights and thresholds in the ANN.

(2) Import the training data into the ANN through data partitioning as a way to obtain the output data.

(3) Calculate the global error of the output data.

(4) Adjust the weights and thresholds.

(5) When the global error is greater than the limit value, the above steps will be repeated. And, when the global error is less than the limit value, the training ends.

2.2.3. Support Vector Regression (SVR)

Support vector machines (SVMs) are affiliated with supervised learning, and they have a strong mathematical foundation and theoretical support. The purpose of SVMs is to find a hyperplane that separates the data in the training set and maximizes the distance from the class domain boundary along the direction perpendicular to the hyperplane, so SVM is also known as the maximum edge algorithm. The SVM algorithm is a supervised ML algorithm derived from the SVM statistical learning theory [30]. SVM has a higher computational accuracy and efficiency for small sample prediction and nonlinear regression analysis than traditional statistical methods [31]. SVM is a supervised learning algorithm primarily used for classification tasks, while SVR is an improvement upon SVM to make it suitable for regression problems. The diagram of the SVM regression structure is shown in Figure 3.
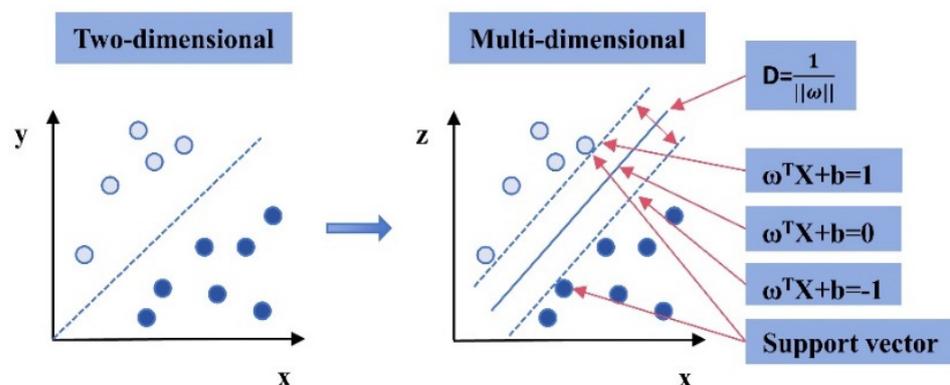
Figure 3. The diagram of the support vector regression model.

The prediction equation for SVR is shown below:

$$f(x) = w^T \cdot \varphi(x) + b \tag{5}$$

where $\varphi(x)$ represents a nonlinear function that maps the input space to the feature space, while w and b, respectively, refer to the weight coefficient and the bias of each feature.

In this work, the brief steps of SVR model training are shown below:

(1) To evaluate the coefficients, it is formulated as a constrained optimization problem [32]:

$$\min\left\{\frac{1}{2} \parallel \omega^2 \parallel + C_p \sum_{n}^{i=1}(\xi_i + \xi_i^*)\right\} \tag{6}$$

$$\begin{aligned} y_i - f(x_i) &\leq \varepsilon + \xi_i \\ f(x_i) - y_i &\geq \varepsilon + \xi_i^* \\ \xi_i \geq 0, \xi_i^* &\geq 0, i = 1, 2, \cdots, n \end{aligned} \tag{7}$$

where C_p is a penalty parameter, ξ_i and ξ_i^* are relaxation factors, and ε is the error tolerance.

(2) Subsequently, an employed Lagrangian function is used to convert the above formula into a dual-optimization problem [33]:

$$\max\left\{\frac{1}{2}\sum_{n}^{i=1}\sum_{n}^{j=1}(a_i - a_i^*)\left(a_j - a_j^*\right)\varphi(x_i)\varphi(x_j) + \varepsilon\sum_{n}^{i=1}(a_i + a_i^*) - \sum_{n}^{i=1}y_i(a_i - a_i^*)\right\} \tag{8}$$

$$\begin{aligned} \sum_{n}^{i=1}(\alpha_i - \alpha_i^*) &= 0 \\ \alpha_i, \alpha_i^* &\in [0, C_p] \end{aligned} \tag{9}$$

where $K(x_i, x_j) = \varphi(x_i)\varphi(x_j)$ represents the kernel function, and α_i and α_i^* are the Lagrangian multipliers.

(3) By solving the above dual-optimization problem, the above formula is expressed as follows:

$$f(x) = \sum_{n}^{i=1}(\alpha_i - \alpha_i^*) \cdot K(x, x_i) + b \tag{10}$$

(4) Due to the exceptional nonlinear characteristics exhibited in the high-dimensional space of radial basis kernel functions (RBF), the following RBF is employed:

$$K(x_i, x_j) = \exp\left\{-k_p \parallel x_i - x_j \parallel^2\right\} \tag{11}$$

where k_p is the kernel parameter.

2.2.4. Random Forests (RFs)

As a statistical learning model, RF extracts multiple sample sets from training samples using self-service resampling techniques, and then constructs decision-tree models using the extracted sample sets. Upon aggregating these decision-tree models, a majority voting or averaging process determines the final result. As a supervised ML algorithm, RF can effectively combine integrated learning and nonlinear statistical methods [34], making the model highly accurate and less prone to overfitting problems. The diagram of the RF regression structure is shown in Figure 4.

The bootstrap aggregating algorithm, often referred to as bagging, is widely utilized for training the RF regression model [35]. In this work, the training set is $T_n = \{(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)\}$, in which X represents the input variables, and Y is the output value. When the training is over, the functional relation $f = (X, T_n)$ can be obtained.

Therefore, we could get m outputs $Y_1^{pre} = f(X, T_n^1), Y_2^{pre} = f(X, T_n^2), \ldots, Y_m^{pre} = f(X, T_n^m)$. After that, the predicted value Y^{pre} is obtained as shown below:

$$Y^{pre} = \frac{1}{m}\sum_{i=1}^{m} Y_i^{pre} = \frac{1}{m}\sum_{i=1}^{m} f(X, T_n^m) \tag{12}$$
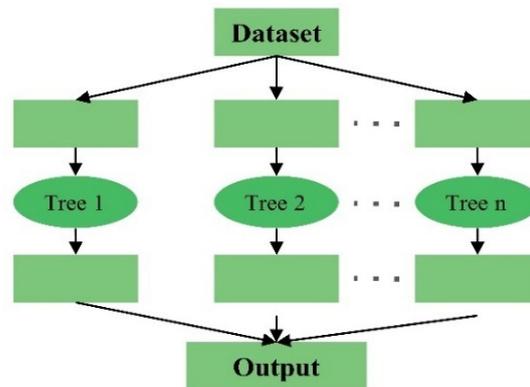
Figure 4. The diagram of the random forests model.

The brief steps of RF model training in this work are shown below:

(1) The fatigue data are collected for use in this study.

(2) The RF regression algorithm selects a suitable split at each node by considering a subset of input variables that are randomly sampled from the available adjustment parameters.

(3) Once the RF model is trained using the collected data, it can be applied to predict fatigue life for new inputs. The trained model utilizes the selected input variables and their corresponding split values to make accurate predictions.

2.3. Model Evaluation

In this work, we utilize two metrics to effectively evaluate the prediction performance. These metrics are R-squared (R^2) and Mean-Squared Error (MSE), and their mathematical expressions are provided below:

$$R^2 = \frac{\sum_{n}^{i=1}(\hat{y}_i - \overline{y}_i)^2}{\sum_{n}^{i=1}(y_i - \overline{y}_i)^2} \tag{13}$$

$$\text{MSE} = \frac{1}{n}\sum_{n}^{i=1}(y_i - \hat{y}_i)^2 \tag{14}$$

where $\hat{y}_i$ is the predicted fatigue life, and y_i is the corresponding experimental fatigue life.

R-squared and MSE are two commonly used metrics for model evaluation in ML. R-squared is a standardized metric that assesses the model's ability to explain sample variability, while MSE is a metric that assesses the size of the average error between the model's predictions and the true value. Both metrics play an important role in assessing model performance and adjusting model parameters. In practice, R-squared and MSE are often used together, with R-squared reflecting the model's fit to the data and MSE reflecting the model's predictive accuracy. First, we need to evaluate the performance of the model based on the R-squared, and then judge the predictive effectiveness of the model in combination with the MSE. If the R-squared is very low, it means that the model cannot explain most of the data variation at this time, even if the MSE is very small, it does not mean that the model is good at prediction.

2.4. Overall Strategy

Figure 5 depicts the flowchart illustrating the machine learning process for predicting fatigue life by considering the synergistic effect of yield stress and fatigue stress. The raw data used in this study were obtained from the article [1]. Prior to applying any machine learning techniques to the collected fatigue test data, data preprocessing is necessary. Based on the analysis of density, porosity, yield stress, and fatigue stress, it has been observed that yield stress and fatigue stress have a significant impact on fatigue life. Hence, four machine learning models were developed to investigate the combined effects of yield stress and fatigue stress on fatigue life. Additionally, to assess the influence of hyperparameters in each machine learning model on prediction accuracy, a discussion on hyperparameters was conducted using R^2 and MSE as evaluation metrics. Based on this analysis, suitable hyperparameter values were recommended for each machine learning model.
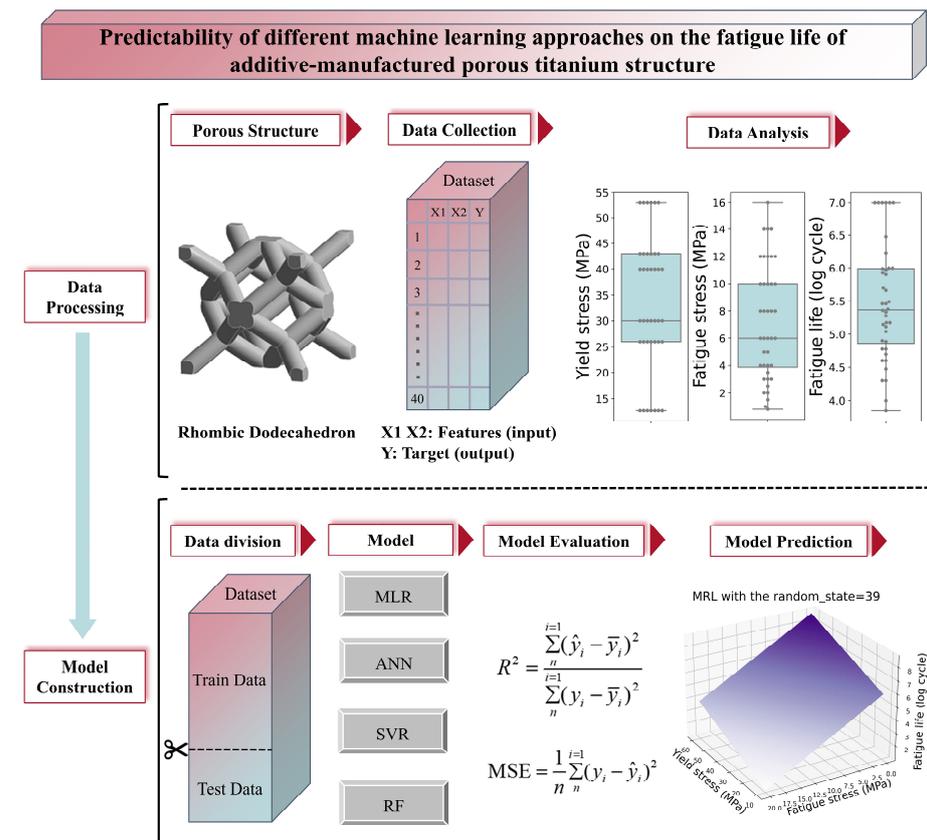
Figure 5. Machine learning process flowchart of predicting fatigue life with the synergic effect of yield stress and fatigue stress.

3. Results

3.1. Data Analysis

The impact of density, porosity, yield stress, and fatigue stress on the fatigue life of the samples is analyzed through data visualization in Figures 6–8. Figure 6 shows the overall 34 sets of experimental data with the fatigue life of the porous samples against the yield stress and the fatigue stress. Figure 7a–e displays the scattered distribution of data points for each variable in the raw data, with the fatigue-life data exhibiting uniform dispersion. Additionally, Figure 7f–g presents the pairwise variations between the variable data, showcasing noticeable correlations between fatigue life and yield stress, as well as fatigue stress, respectively. The correlation coefficients provided in Figure 8 further confirm these relationships, particularly highlighting the significant influence of fatigue stress on fatigue life. To emphasize the importance of data quality on the predictive performance of the ML model, Figure 6 visualizes the raw data distributions of yield stress and fatigue

stress in relation to fatigue life. Notably, yield stress demonstrates a positive correlation with fatigue life, while fatigue stress exhibits a negative correlation with fatigue life.
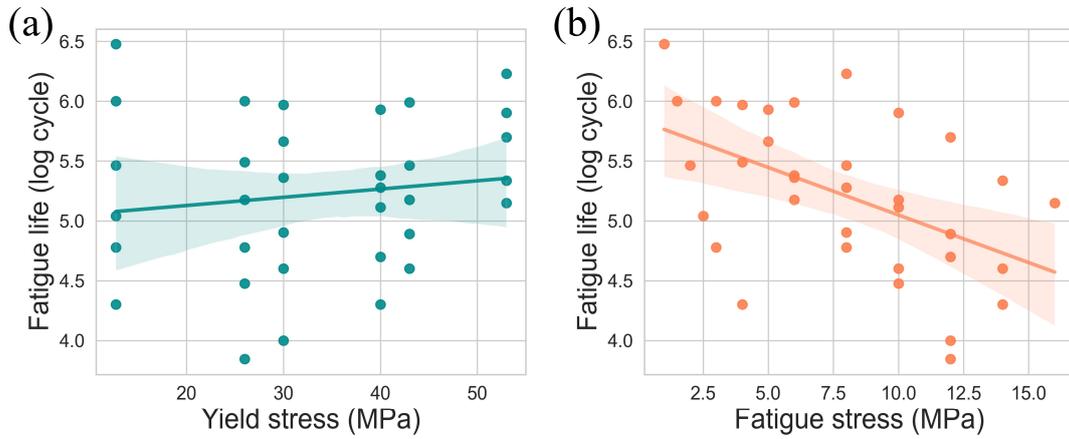
Figure 6. The 34 sets of experimental data with the fatigue life of the porous samples against (**a**) the yield stress and (**b**) the fatigue stress.
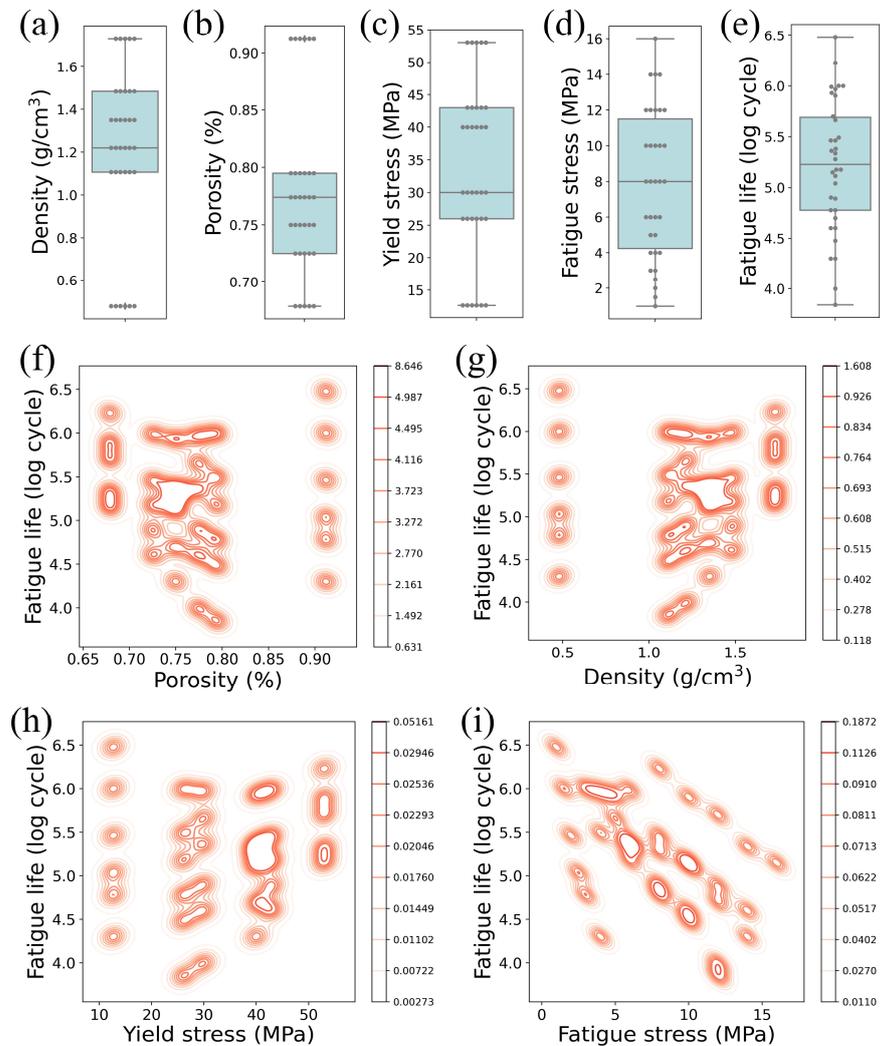
Figure 7. Visualization of experimental data: (**a**–**e**) box line plots of density, porosity, yield stress, fatigue stress, and fatigue life; (**f**–**i**) density plots of density, porosity, yield stress, and fatigue stress against fatigue life, respectively.
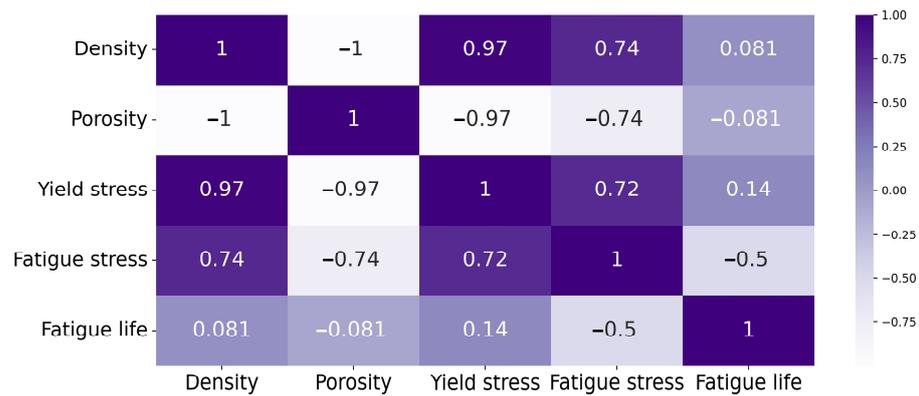
Figure 8. Heatmaps of density, porosity, yield stress, fatigue stress, and fatigue life.

3.2. Fatigue-Life Prediction

This study employs four ML models, namely MLR, ANN, SVR, and RF, to predict the fatigue life of AM titanium porous components. The training set constitutes 80% of the total data, while the remaining 20% is allocated for testing purposes.

For the MLR model, a comparison of the predicted fatigue life data and the experimental fatigue life is shown in Figure 9a–c. These circles indicate the comparison of the predicted values with the experimental values, which are ideally equal, i.e., they correspond to the black line in the figure. Thus, an initial visual indicator of model performance is the proximity of these circles to this ideal black line. We can find that each circle is scattered around the desired black line in the training set, test set, and all the data, indicating a better prediction performance of the MLR model. Figure 9d illustrates the 95% confidence intervals, demonstrating that all the data points are also dispersed within the upper and lower dashed lines. This observation provides supporting evidence for the notion that the MLR model exhibits superior predictive performance.
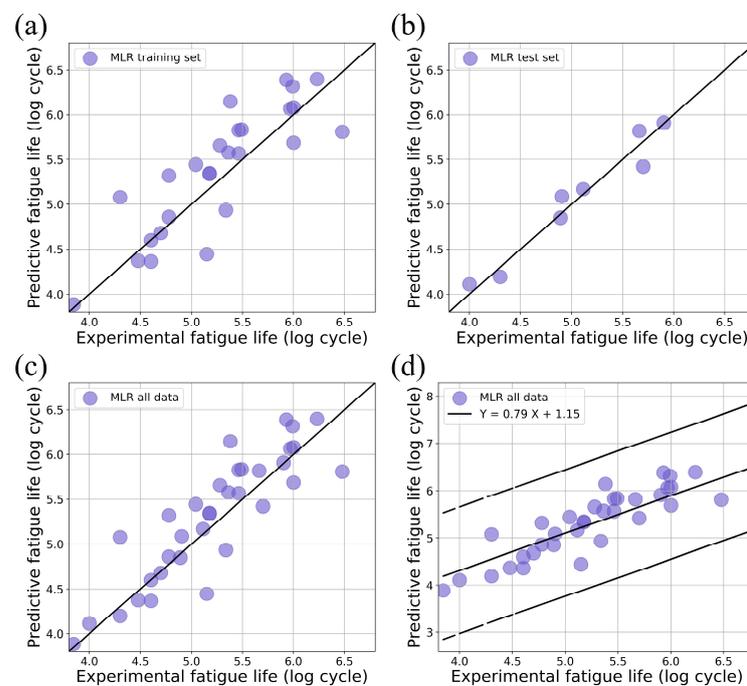
Figure 9. Comparison of MLR predictive fatigue life and experimental fatigue life: (**a**) training set, (**b**) test set, (**c**) all data, and (**d**) 95% confidence interval.

Figures 10–12 present a comparison between the predicted fatigue life data and the corresponding experimental fatigue life for the ANN, SVR and RF models, respectively.

As observed in Figures 10d, 11d and 12d, the prediction performance of each model corresponds to a 95% confidence interval, indicating their suitability for predicting fatigue life. However, upon closer inspection of Figures 10–12, it becomes apparent that the SVR model exhibits superior predictive performance among the three models, as evidenced by the close proximity of the circles to the ideal black line.
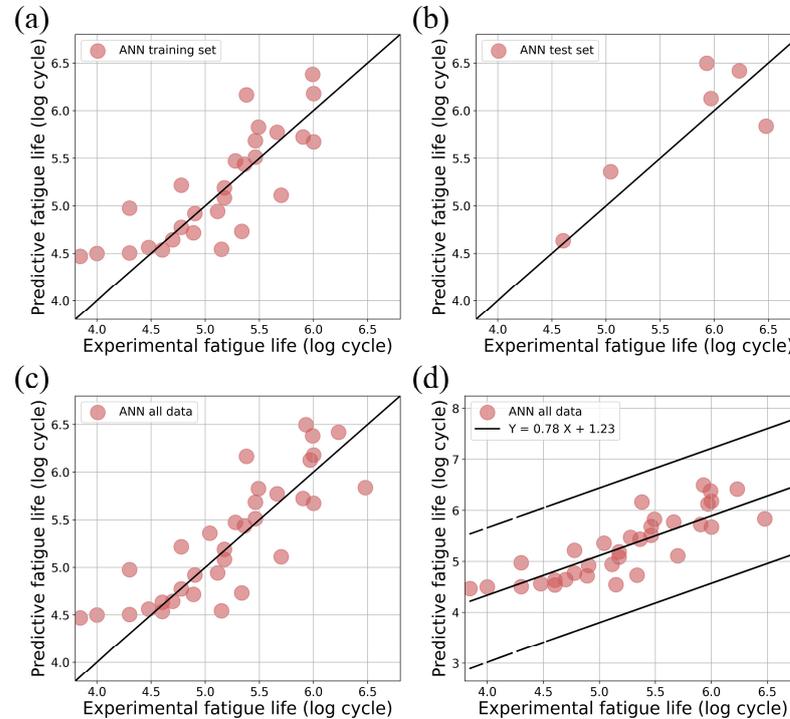
Figure 10. Comparison of ANN predictive fatigue life and experimental fatigue life: (**a**) training set, (**b**) test set, (**c**) all data, and (**d**) 95% confidence interval.
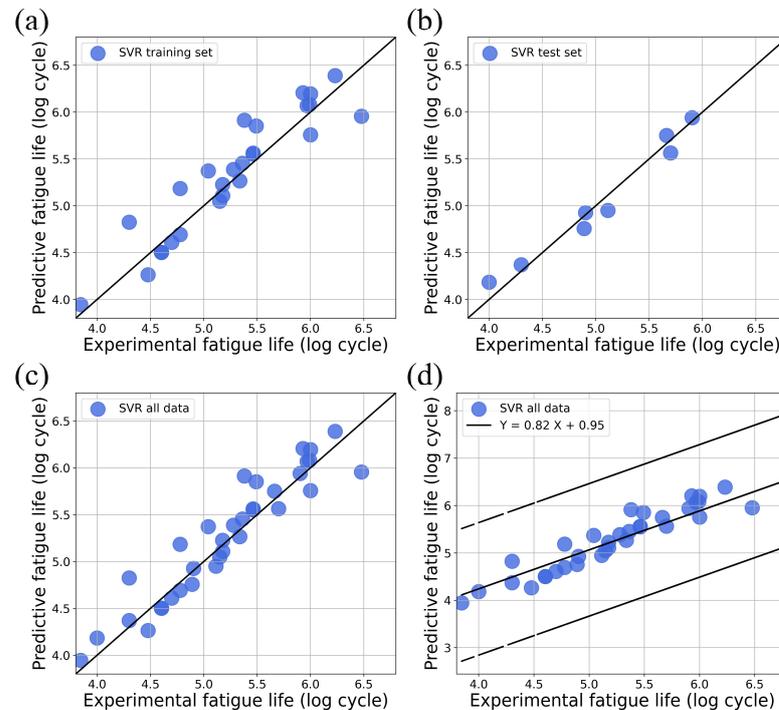
Figure 11. Comparison of SVR predictive fatigue life and experimental fatigue life: (**a**) training set, (**b**) test set, (**c**) all data, and (**d**) 95% confidence interval.
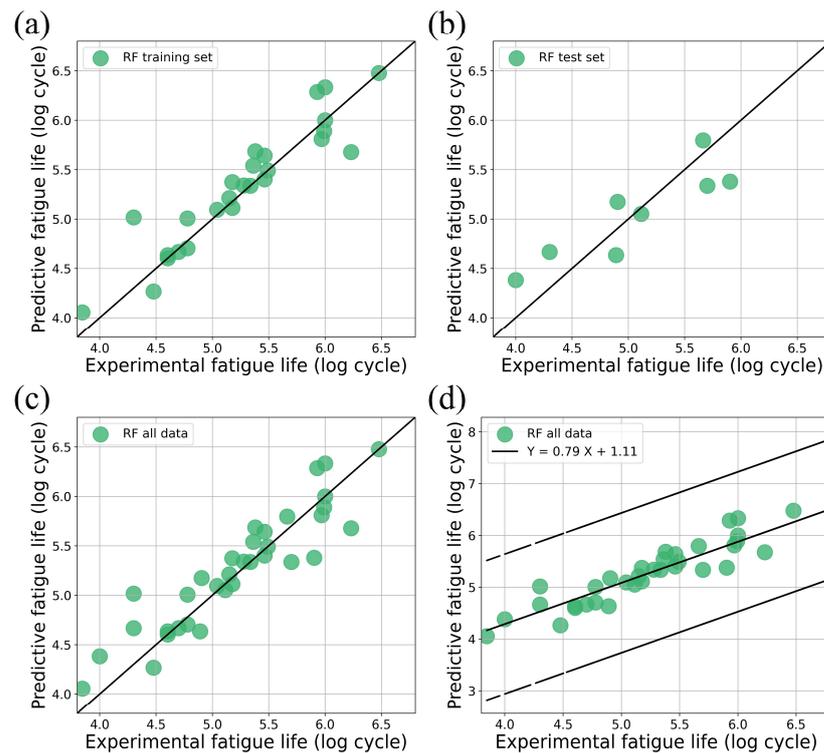
Figure 12. Comparison of RF predictive fatigue life and experimental fatigue life: (**a**) training set, (**b**) test set, (**c**) all data, and (**d**) 95% confidence interval.

The analysis of the results and the comparison between the predicted and experimental data for AM titanium porous components indicate that the proposed method is highly effective in predicting fatigue life.

3.3. Performance of the Models

This study assesses the prediction performance of MLR, ANN, SVR and RF models for AM titanium porous components. It is noteworthy that all four models demonstrate strong predictive capabilities in both the training and test sets, as well as across all data. Furthermore, the predictive power of each model falls within a 95% confidence interval.

4. Discussion

4.1. Effects of MLR Parameters on Predicted Results and Prediction Accuracy

In this subsection, we discuss in detail the impact of the different training sets and test sets on the predicted fatigue life of AM titanium porous components. This is mainly reflected in the use of unused random seeds when dividing the data.

Machine learning models are sensitive to the quality and characteristics of the data they are trained on. The performance and generalization ability of MLR models heavily rely on the data that is used for training. Table 1 presents the various training sets utilized for the MLR models in this study. To facilitate a fair comparison of predicted fatigue life, other hyperparameters were kept constant during the training process. Figure 13 presents an analysis of the fatigue-life prediction by four MLR models, revealing consistent predictive trends across all models. Importantly, training the MLR model using different training sets has minimal impact on the predictive trend of the model.

In addition, Figure 14 displays the performance evaluation (R-squared and MSE) for the MLR model on a training and test dataset. The R-square of the MLR model gradually increases on the training set and decreases on the test set as random states increase (Figure 14a,b), while the R-squared and MSE of the MLR model on all data show a constant trend with the increase in random states (Figure 14c).

Table 1. Different MLR models for fatigue-life prediction of AM titanium porous components.

MLR	Hyperparameters
Model 1	random_states = 39
Model 2	random_states = 50
Model 3	random_states = 74
Model 4	random_states = 110
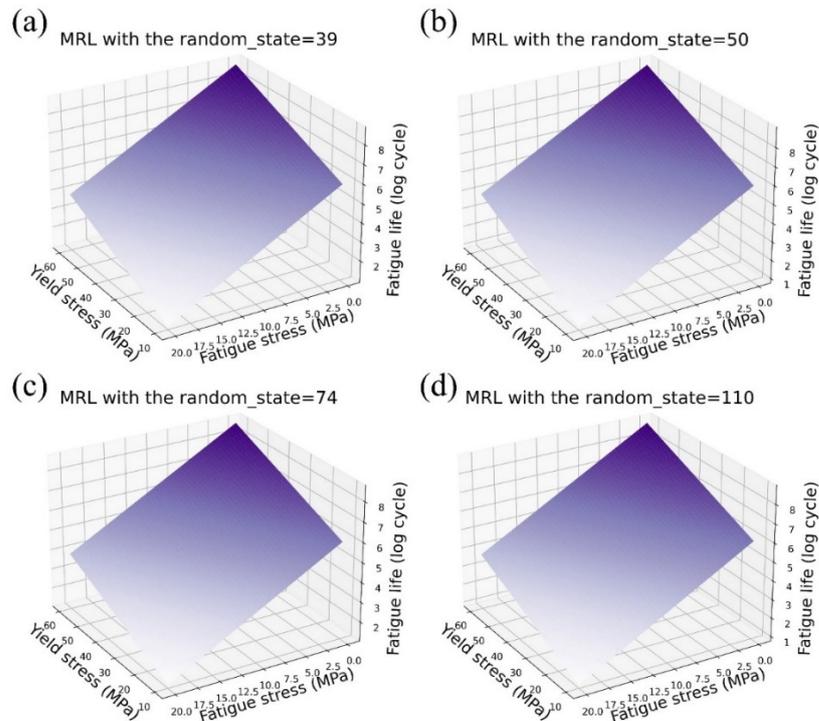
Figure 13. MLR visualization with different random states: (**a**) 39, (**b**) 50, (**c**) 74, and (**d**) 110.
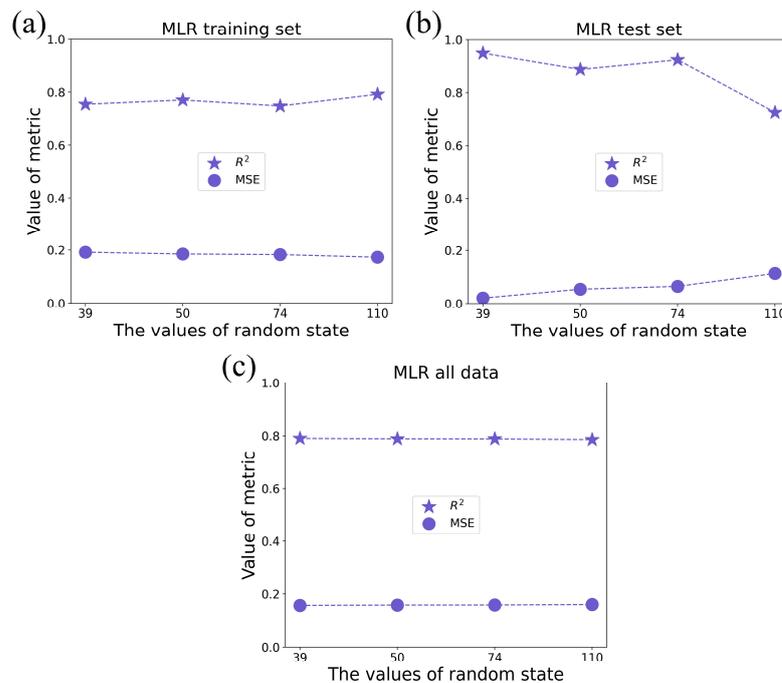
Figure 14. Visualization of MLR prediction accuracy with different hyperparameters: (**a**) training, set, (**b**) test set, and (**c**) all data.

In the investigation of hyperparameters for the MLR model used in predicting fatigue life of AM titanium porous components, it was observed that variations in random states have minimal impact on the predictive performance of the MLR model. Below are the four MLR models considered in this study:

$$\text{Model 1}: N = 5.3320656 + 0.0569881\sigma_y - 0.24416625\sigma_f \tag{15}$$

$$\text{Model 2}: N = 5.3049312 + 0.0597426\sigma_y - 0.2488433\sigma_f \tag{16}$$

$$\text{Model 3}: N = 5.3965109 + 0.0541417\sigma_y - 0.2381241\sigma_f \tag{17}$$

$$\text{Model 4}: N = 5.4335601 + 0.0566145\sigma_y - 0.2541425\sigma_f \tag{18}$$

where N is the fatigue life (number of cycles to failure), σ_y is the yield stress, and σ_f is the fatigue stress.

4.2. Effects of ANN Parameters on Predicted Results and Prediction Accuracy

In general, there is no universally applicable rule for determining the optimal number of hidden layers and neurons in ANN models. Therefore, in this subsection, we discuss in detail the impact of the number of neurons in the first hidden layer on the predicted fatigue life of AM titanium porous components.

Undoubtedly, the number of neurons in the first hidden layer plays a crucial role in determining the predictive performance of an ANN model. Table 2 presents the number of neurons in the first hidden layer for the ANN models used in this study. To ensure an accurate comparison of fatigue lives predicted by different models, other hyperparameters are kept constant during the training process. As depicted in Figure 15, each of the four ANN models exhibits a consistent predictive trend for fatigue life. However, it is evident from Figure 15c,d that increasing the number of neurons in the first hidden layer gradually causes the ANN model to become locally overfitted and decrease its generalization ability. Conversely, as shown in Figure 15a,b, when the number of neurons in the first hidden layer is three or four, the prediction surface of the ANN model remains smooth without any signs of local overfitting.

Furthermore, the performance of the ANN model (measured by R-squared and MSE) on the training, test, and overall datasets is illustrated in Figure 16. As depicted in Figure 16a,b, it can be observed that increasing the number of neurons in the first hidden layer of the ANN model leads to a gradual increase in the R-squared value on the training set. However, the R-squared on the test set initially increases and then decreases, indicating a decrease in the generalization ability of the ANN model, which aligns with the observations from Figure 15. Figure 16c demonstrates that as the number of neurons in the first hidden layer increases, the R-squared on all data gradually improves while the MSE steadily decreases.

According to our assessment of the hyperparameters of the ANN model, three or four neurons in the first hidden layer of the ANN model are most suitable for the fatigue-life prediction of AM titanium porous components. An ANN model with a smaller MSE or a larger R-squared does not necessarily indicate more predictive ability. In addition to R-squared and MSE, the predictive trend of an ANN model should also be taken into account when evaluating its predictive ability. The purpose of this is to prevent local overfitting during ANN model training and maximize generalization ability.

Table 2. Different ANN models for fatigue-life prediction of AM titanium porous components.

ANN	Hyperparameters
Model 1	The first hidden layer has 3 neurons
Model 2	The first hidden layer has 4 neurons
Model 3	The first hidden layer has 5 neurons
Model 4	The first hidden layer has 6 neurons
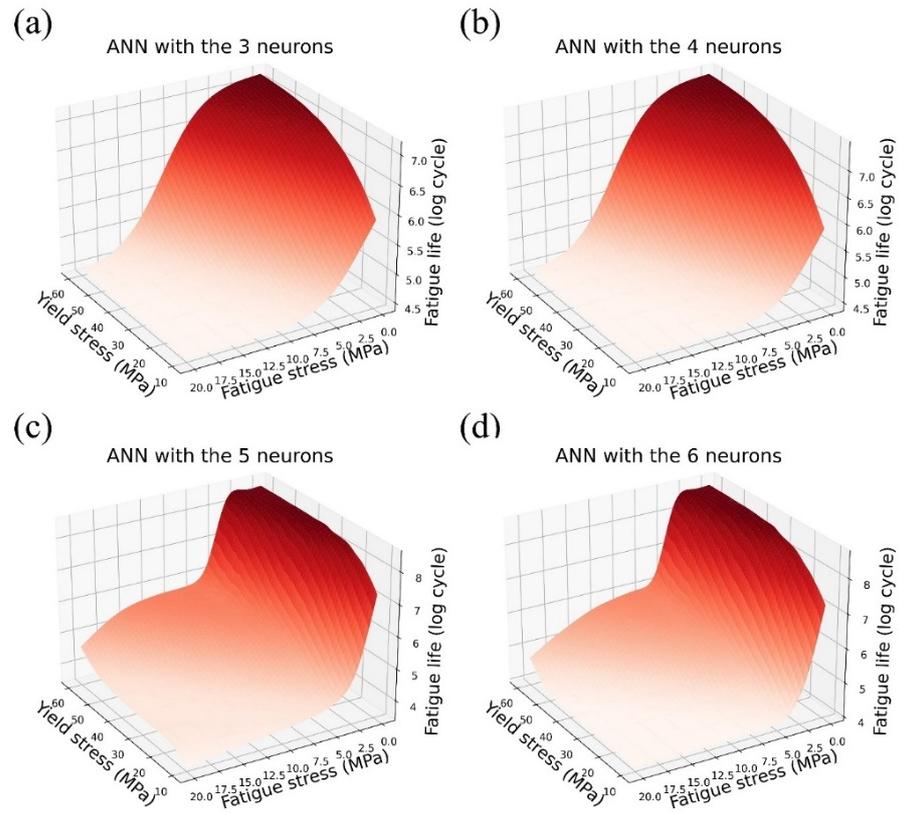
Figure 15. ANN visualization with different neurons in the first hidden layer: (**a**) 3, (**b**) 4, (**c**) 5, and (**d**) 6.
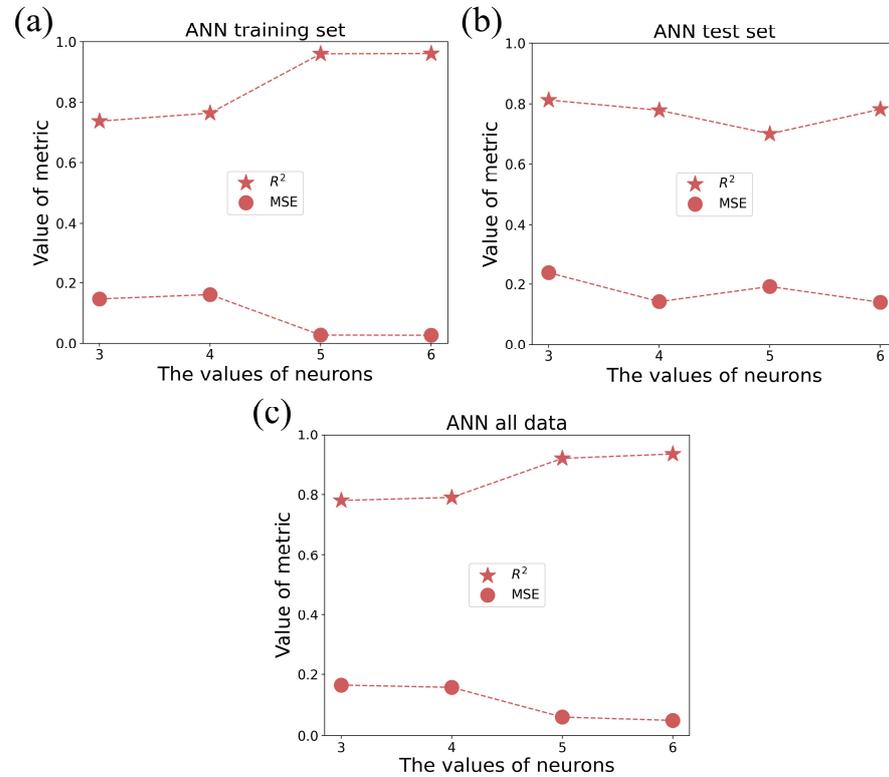
Figure 16. Visualization of ANN prediction accuracy with different hyperparameters: (**a**) training set, (**b**) test set, and (**c**) all data.

4.3. Effects of SVR Parameters on Predicted Results and Prediction Accuracy

The performance and complexity of an SVR model are primarily determined by the parameters gamma and C. Hence, in this section, we provide a comprehensive analysis of how gamma and C influence the predicted fatigue life of AM titanium porous components.

To achieve better performance of the SVR model, optimal gamma and C values can be selected by training and evaluating on different combinations of parameter values. In this study, the hyperparameters of the SVR model are presented in Table 3, and all other hyperparameters are kept constant during the SVR model training process to ensure the fair comparison of the fatigue-life prediction between different models. As shown in Figure 17, each SVR model exhibits a consistent trend in predicting fatigue life. However, as illustrated in Figure 17a–c, increasing C gradually and setting gamma to 0.001 leads to local overfitting of the SVR model and decreases its generalization ability. Similarly, Figure 17c,d demonstrates that with the gradual increase in gamma and C equal to 30, the SVR model becomes more and more locally overfit, decreasing its generalization abilities. Additionally, Figure 18 shows the performance evaluation of the SVR model (R-squared and MSE) on the training, test, and total datasets. As shown in Figure 18a,b, increasing C and setting gamma to 0.001 leads to a gradual increase in R-squared on the training set, while the R-squared on the test set decreases gradually. From Figure 18a,b, on the training set, the R-square of the SVR model gradually increases with gamma, and on the test set, the R-square first decreases and then increases. From Figure 18c, as gamma and C increase, the R-squared of the SVR model shows an increasing–decreasing–increasing trend. On the other hand, the MSE shows a decreasing–increasing–decreasing trend.

Based on the study of hyperparameters, the SVR model with gamma equal to 0.0001 and C equal to 30 is determined to be the most suitable for predicting the fatigue life of AM titanium porous components. It is important to note that there is no direct correlation between a smaller MSE and a higher R-squared and the predictive performance of the SVR model. Therefore, it is crucial to consider the predictive trend of the SVR model when evaluating its predictive ability, not just the R-squared and MSE values.

Table 3. Different SVR models for fatigue-life prediction of AM titanium porous components.

SVR	Hyperparameters
Model 1	gamma = 0.001 and C = 10
Model 2	gamma = 0.001 and C = 50
Model 3	gamma = 0.001 and C = 416
Model 4	gamma = 0.0001 and C = 30
Model 5	gamma = 0.005 and C = 30
Model 6	gamma = 0.01 and C = 30

4.4. Effects of RF Parameters on Predicted Results and Prediction Accuracy

In general, the parameters n_estimators and max_depth are typically used in RF models to control model complexity and performance. Therefore, in this subsection, we discuss in detail the impact of the n_estimators and max_depth.

To determine the best values, these parameters must usually be tuned and experimented with according to the dataset, the difficulty of the problem, and the available computational resources. In this work, the hyperparameters of the RF model are set as shown in Table 4. The other hyperparameters are maintained as constant during the training process of the RF model in order to better compare the predicted fatigue life of different models. Figure 19 shows a visualization of the fatigue life predicted by the six RF models, observing a consistent trend in the predictions for each model. However, from Figure 19a–c, by gradually increasing max_depth and n_estimators equal to three, the RF model's local overfitting is reduced and its generalization ability is enhanced. From Figure 19d–f, the RF model's generalization capability increases as the max_depth equals to five and n_estimators are gradually increased.
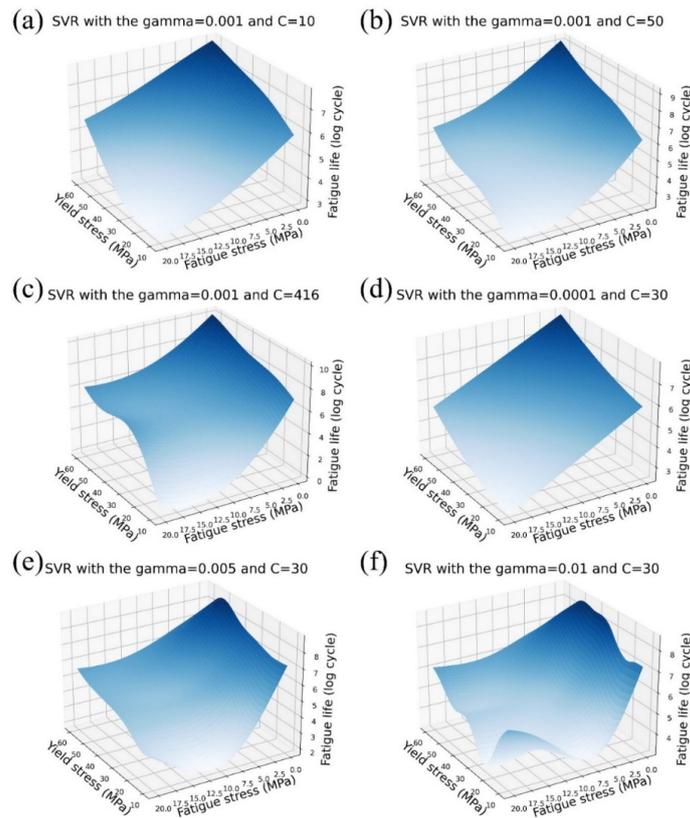
Figure 17. SVR visualization with different gamma and C: (**a**) 0.001 and 10, (**b**) 0.001 and 50, (**c**) 0.001 and 416, (**d**) 0.0001 and 30, (**e**) 0.005 and 30, and (**f**) 0.001 and 30.
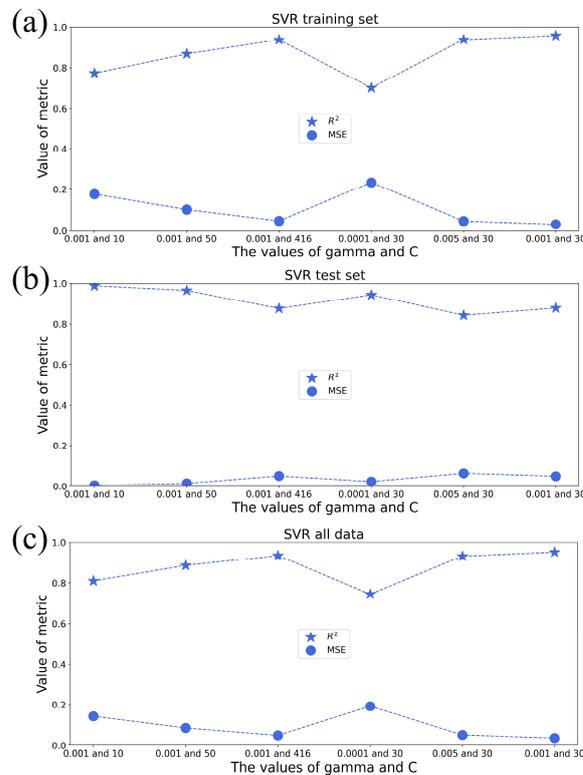
Figure 18. Visualization of SVR prediction accuracy with different hyperparameters: (**a**) training set, (**b**) test set, and (**c**) all data.

Table 4. Different RF models for fatigue-life prediction of AM titanium porous components.

RF	Hyperparameters
Model 1	n_estimators = 3 and max_depth = 3
Model 2	n_estimators = 3 and max_depth = 5
Model 3	n_estimators = 3 and max_depth = 7
Model 4	n_estimators = 5 and max_depth = 5
Model 5	n_estimators = 7 and max_depth = 5
Model 6	n_estimators = 9 and max_depth = 5
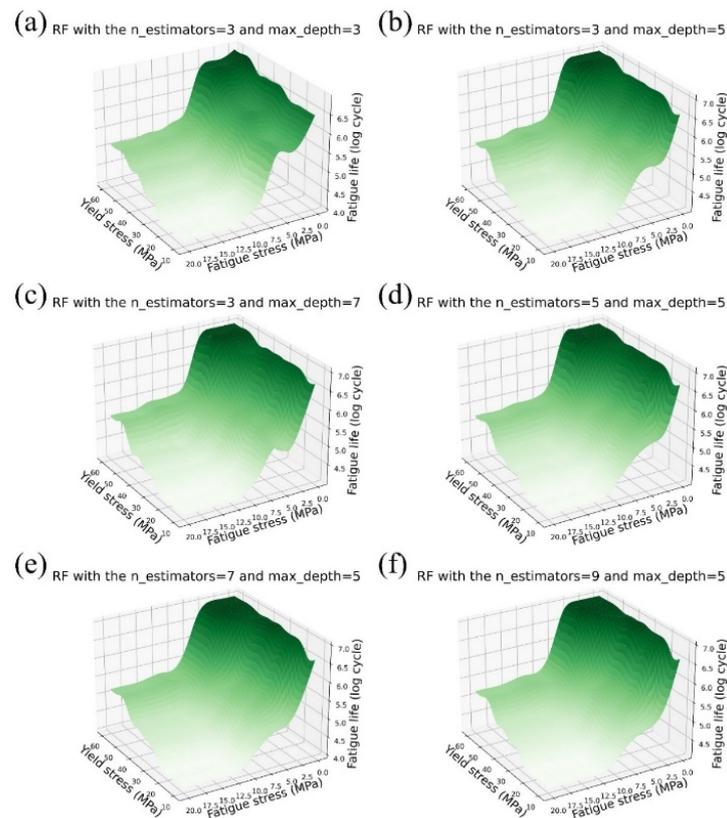
Figure 19. RF visualization with different n_estimators and max_depth: (**a**) 3 and 3, (**b**) 3 and 5, (**c**) 3 and 7, (**d**) 5 and 5, (**e**) 7 and 5, and (**f**) 9 and 5.

Furthermore, Figure 20 displays the performance evaluation (R^2 and MSE) of the RF model on the training set, test set, and overall data. As illustrated in Figure 20a,b, when n_estimators is set to three and max_depth is gradually increased, the R-squared of the RF model gradually increases on both the training set and the test set. Additionally, Figure 20a,b indicates that when max_depth is set to 5 and n_estimators is gradually increased, the R-squared on the training set gradually increases while the R-squared on the test set gradually decreases. Moreover, Figure 20c shows that as both max_depth and n_estimators increase, the R-squared of the RF model on the overall data exhibits a gradual increase, while the MSE decreases.

Based on the study of the hyperparameters of the RF model for the fatigue-life prediction of AM titanium porous components, it is found that the RF model with n_estimators equal to three and max_depth equal to seven is the most suitable for fatigue-life prediction. It is worth noting that neither a smaller MSE nor a larger R-squared indicates the predictive ability of the RF model. Evaluating the predictive ability of an RF model should not only look at the R-squared and MSE, but also be combined with the predictive trend of the model. This is to avoid the occurrence of local overfitting during RF model training and to maximize the generalization ability of the model.
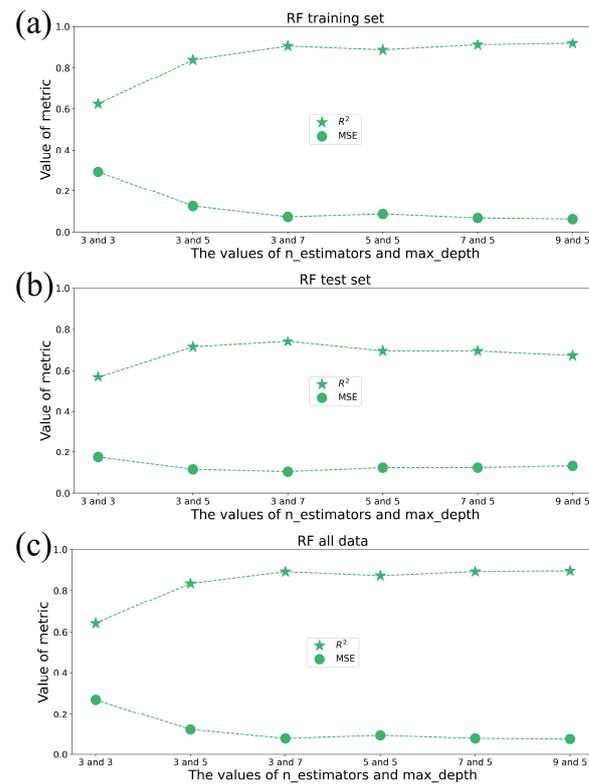
Figure 20. Visualization of RF prediction accuracy with different hyperparameters: (**a**) training set, (**b**) test set, and (**c**) all data.

4.5. Remarks

Indeed, the rapid development of big data and artificial intelligence has ushered in the fourth paradigm of materials science research. Machine learning plays a crucial role in enhancing researchers' reliance on intuition and reducing the need for extensive trial and error. This leads to cost savings in experiments and allows for a thorough exploration of potential connections within experimental data, ultimately shaping a data-driven research model.

As illustrated in Figure 15, we propose that the first hidden layer of the artificial neural network model should consist of three or four neurons for predicting the fatigue life of the additively manufactured titanium porous component. This recommendation stems from the observed smoothness of the predicted surface, which indicates an absence of fitting anomalies. The discourse on hyperparameters in the other three models serves as a strategic approach to address the overfitting issue, not only within the artificial neural network but also across the broader scope of our research.

In this work, the ML-based method was used to predict the fatigue life of AM titanium. However, it is important to acknowledge that other factors such as microstructure and various process parameters [36–40] can also potentially influence the fatigue life. To further enhance the accuracy and robustness of the prediction model, it is recommended that future research considers incorporating as many influencing variables as possible. By including additional factors into the analysis, a more comprehensive understanding of the relationship between these variables and fatigue life can be achieved, leading to improved predictions and insights in the field of AM titanium fatigue life assessment.

The integration of machine learning models with physical knowledge is also crucial. By incorporating physical knowledge into machine learning models through algorithm learning, sample output, and data observation, we can not only address the "black box" issue in machine learning models but also enhance the transparency, interpretability, and prediction accuracy of these models.

Finally, due to the limitation of data availability, the present work relies on experimental data. Further validation, possibly through cross-validation or external datasets, will be performed in future studies.

5. Conclusions

In this work, four machine learning models were used, MLR, ANN, SVR, and RF, to predict the fatigue life of AM titanium porous components. Key findings are summarized as follows:

1. The MLR model's predictions of fatigue life for AM titanium porous components are not significantly affected by variations in the training sets used.
2. To achieve accurate predictions of fatigue life for AM titanium porous components using the ANN model, it is recommended to create the first hidden layer with three or four neurons.
3. For the SVR model, gamma equal to 0.0001 and C equal to 30 are recommended for the fatigue-life prediction of AM titanium porous components.
4. For accurate predictions of fatigue life in AM titanium porous components using the RF model, it is suggested to set the n_estimators equal to three and the max_depth equal to seven.

The primary conclusions can also serve as a valuable reference for future researchers aiming to predict the fatigue life of porous titanium alloy components manufactured through additive manufacturing techniques.

Author Contributions: S.G.: methodology, formal analysis, investigation, data curation, writing—original draft preparation, writing—review and editing; X.Y.: methodology, validation, writing—review and editing; H.W.: conceptualization, methodology, validation, writing—review and editing, project administration. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Liu, Y.; Wang, H.; Li, S.; Wang, S.; Wang, W.; Hou, W.; Hao, Y.; Yang, R.; Zhang, L. Compressive and fatigue behavior of beta-type titanium porous structures fabricated by electron beam melting. *Acta Mater.* **2017**, *126*, 58–66. [CrossRef]
2. Li, S.; Murr, L.; Cheng, X.; Zhang, Z.; Hao, Y.; Yang, R.; Medina, F.; Wicker, R. Compression fatigue behavior of Ti–6Al–4V mesh arrays fabricated by electron beam melting. *Acta Mater.* **2012**, *60*, 793–802. [CrossRef]
3. Zhang, L.C.; Klemm, D.; Eckert, J.; Hao, Y.L.; Sercombe, T.B. Manufacture by selective laser melting and mechanical behavior of a biomedical Ti–24Nb–4Zr–8Sn alloy. *Scr. Mater.* **2011**, *65*, 21–24. [CrossRef]
4. Liu, Y.; Li, S.; Wang, H.; Hou, W.; Hao, Y.; Yang, R.; Sercombe, T.; Zhang, L. Microstructure, defects and mechanical behavior of beta-type titanium porous structures manufactured by electron beam melting and selective laser melting. *Acta Mater.* **2016**, *113*, 56–67. [CrossRef]
5. Zhao, S.; Li, S.; Hou, W.; Hao, Y.; Yang, R.; Misra, R. The influence of cell morphology on the compressive fatigue behavior of Ti-6Al-4V meshes fabricated by electron beam melting. *J. Mech. Behav. Biomed. Mater.* **2016**, *59*, 251–264. [CrossRef]
6. Edwards, P.; O'Conner, A.; Ramulu, M. Electron Beam Additive Manufacturing of Titanium Components: Properties and Performance. *J. Manuf. Sci. Eng.* **2013**, *135*, 61016. [CrossRef]
7. Zadpoor, A.A. Bone tissue regeneration: The role of scaffold geometry. *Biomater. Sci.* **2014**, *3*, 231–245. [CrossRef]
8. Lietaert, K.; Cutolo, A.; Boey, D.; Van Hooreweder, B. Fatigue life of additively manufactured Ti6Al4V scaffolds under tension-tension, tension-compression and compression-compression fatigue load. *Sci. Rep.* **2018**, *8*, 4957. [CrossRef] [PubMed]

9. Wycisk, E.; Emmelmann, C.; Siddique, S.; Walther, F. High cycle fatigue (HCF) performance of Ti-6Al-4V alloy processed by selective laser melting. In *Advanced Materials Research*; Trans Tech Publications Ltd.: Zürich, Switzerland, 2013; pp. 134–139.

10. Hrabe, N.W.; Heinl, P.; Flinn, B.; Körner, C.; Bordia, R.K. Compression-compression fatigue of selective electron beam melted cellular titanium (Ti-6Al-4V). *J. Biomed. Mater. Res. Part B Appl. Biomater.* **2011**, *99*, 313–320. [CrossRef] [PubMed]

11. Jamshidinia, M.; Wang, L.; Tong, W.; Ajlouni, R.; Kovacevic, R. Fatigue properties of a dental implant produced by electron beam melting ®(EBM). *J. Am. Acad. Dermatol.* **2015**, *226*, 255–263. [CrossRef]

12. Lindemann, J.; Wagner, L. Mean stress sensitivity in fatigue of α, (αβ) and β titanium alloys. *Mater. Sci. Eng. A* **1997**, *234*, 1118–1121. [CrossRef]

13. Henry, S.D.; Dragolich, K.S.; DiMatteo, N. *Fatigue Data Book: Light Structural Alloys*; ASM International: Novelty, OH, USA, 1995.

14. Nicholas, T. *High Cycle Fatigue: A Mechanics of Materials Perspective*; Elsevier: Amsterdam, The Netherlands, 2006.

15. Van Hooreweder, B.; Apers, Y.; Lietaert, K.; Kruth, J.-P. Improving the fatigue performance of porous metallic biomaterials produced by Selective Laser Melting. *Acta Biomater.* **2017**, *47*, 193–202. [CrossRef]

16. Ye, S.; Li, B.; Li, Q.; Zhao, H.-P.; Feng, X.-Q. Deep neural network method for predicting the mechanical properties of composites. *Appl. Phys. Lett.* **2019**, *115*, 161901. [CrossRef]

17. Naik, D.L.; Kiran, R. Identification and characterization of fracture in metals using machine learning based texture recognition algorithms. *Eng. Fract. Mech.* **2019**, *219*, 6618. [CrossRef]

18. Ma, X.; He, X.; Tu, Z. Prediction of fatigue–crack growth with neural network-based increment learning scheme. *Eng. Fract. Mech.* **2020**, *241*, 107402. [CrossRef]

19. Zhang, M.; Sun, C.-N.; Zhang, X.; Goh, P.C.; Wei, J.; Hardacre, D.; Li, H. High cycle fatigue life prediction of laser additive manufactured stainless steel: A machine learning approach. *Int. J. Fatigue* **2019**, *128*, 105194. [CrossRef]

20. Zhan, Z.; Li, H. Machine learning based fatigue life prediction with effects of additive manufacturing process parameters for printed SS 316L. *Int. J. Fatigue* **2020**, *142*, 105941. [CrossRef]

21. Zhan, Z.; Li, H. A novel approach based on the elastoplastic fatigue damage and machine learning models for life prediction of aerospace alloy parts fabricated by additive manufacturing. *Int. J. Fatigue* **2020**, *145*, 106089. [CrossRef]

22. Zhan, Z.; Ao, N.; Hu, Y.; Liu, C. Defect-induced fatigue scattering and assessment of additively manufactured 300M-AerMet100 steel: An investigation based on experiments and machine learning. *Eng. Fract. Mech.* **2022**, *264*, 108352. [CrossRef]

23. Shi, T.; Sun, J.; Li, J.; Qian, G.; Hong, Y. Machine learning based very-high-cycle fatigue life prediction of AlSi10Mg alloy fabricated by selective laser melting. *Int. J. Fatigue* **2023**, *171*, 7585. [CrossRef]

24. Bao, H.; Wu, S.; Wu, Z.; Kang, G.; Peng, X.; Withers, P.J. A machine-learning fatigue life prediction approach of additively manufactured metals. *Eng. Fract. Mech.* **2020**, *242*, 107508. [CrossRef]

25. Romano, S.; Brandão, A.; Gumpinger, J.; Gschweitl, M.; Beretta, S. Qualification of AM parts: Extreme value statistics applied to tomographic measurements. *Mater. Des.* **2017**, *131*, 32–48. [CrossRef]

26. du Plessis, A.; Yadroitsava, I.; Yadroitsev, I. Effects of defects on mechanical properties in metal additive manufacturing: A review focusing on X-ray tomography insights. *Mater. Des.* **2019**, *187*, 108385. [CrossRef]

27. Srinivasan, V.; Valsan, M.; Rao, K.B.S.; Mannan, S.; Raj, B. Low cycle fatigue and creep–fatigue interaction behavior of 316L(N) stainless steel and life prediction by artificial neural network approach. *Int. J. Fatigue* **2003**, *25*, 1327–1338. [CrossRef]

28. Nasiri, S.; Khosravani, M.R.; Weinberg, K. Fracture mechanics and mechanical fault detection by artificial intelligence methods: A review. *Eng. Fail. Anal.* **2017**, *81*, 270–293. [CrossRef]

29. Apicella, A.; Donnarumma, F.; Isgrò, F.; Prevete, R. A survey on modern trainable activation functions. *Neural Netw.* **2021**, *138*, 14–32. [CrossRef] [PubMed]

30. Noble, W.S. What is a support vector machine? *Nat. Biotechnol.* **2006**, *24*, 1565–1567. [CrossRef]

31. Wang, L.; Zhu, S.-P.; Luo, C.; Liao, D.; Wang, Q. Physics-guided machine learning frameworks for fatigue life prediction of AM materials. *Int. J. Fatigue* **2023**, *172*, 7658. [CrossRef]

32. Kang, S.; Cho, S. Approximating support vector machine with artificial neural network for fast prediction. *Expert Syst. Appl.* **2014**, *41*, 4989–4995. [CrossRef]

33. French, M. *Fundamentals of Optimization*; Springer: Cham, Switzerland, 2018.

34. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]

35. Zhang, D.; Zhou, X.; Leung, S.C.H.; Zheng, J. Vertical bagging decision trees model for credit scoring. *Expert Syst. Ap-Plications* **2010**, *37*, 7838–7843. [CrossRef]

36. Konda, N.; Verma, R.; Jayaganthan, R. Machine Learning Based Predictions of Fatigue Crack Growth Rate of Additively Man-ufactured Ti6Al4V. *Metals* **2021**, *12*, 50. [CrossRef]

37. Yuan, M.; Zhao, X.; Yue, Q.; Gu, Y.; Zhang, Z. The Effect of Microstructure on the Very High Cycle Fatigue Behavior of Ti-6Al-4V Alloy. *Metals* **2024**, *14*, 254. [CrossRef]

38. Konecna, R.; Varmus, T.; Nicoletto, G.; Jambor, M. Influence of Build Orientation on Surface Roughness and Fatigue Life of the Al2024-RAM2 Alloy Produced by Laser Powder Bed Fusion (L-PBF). *Metals* **2023**, *13*, 1615. [CrossRef]

39.   Spignoli, N.; Minak, G. Influence on Fatigue Strength of Post-Process Treatments on Thin-Walled AlSi10Mg Structures Made by Additive Manufacturing. *Metals* **2023**, *13*, 126. [CrossRef]

40.   Martins, L.F.L.; Provencher, P.R.; Brochu, M.; Brochu, M. Effect of Platform Temperature and Post-Processing Heat Treatment on the Fatigue Life of Additively Manufactured AlSi7Mg Alloy. *Metals* **2021**, *11*, 679. [CrossRef]