ACCSCIENCE PUBLISHING

International Journal of AI for Materials and Design

ORIGINAL RESEARCH ARTICLE

Improvement of multiaxial fatigue life prediction performance based on contrastive learning feature extraction

Ziyu Cui[1], Xingyue Sun[2*], and Xu Chen[1,3]

[1]Department of Process Equipment & Control Engineering, School of Chemical Engineering and Technology, Tianjin University, Tianjin, China

[2]Department of Aeronautical Structure Engineering, School of Aeronautics, Northwestern Polytechnical University, Xi'an, Shaanxi, China

[3]Zhejiang Institute of Tianjin University, Ningbo, Zhejiang, China

(This article belongs to the *Special Issue: AI for Multiscale Analysis and Defect Identification in Packaging Structures and Semiconductor Chips*)

Abstract

Accurate prediction of multiaxial fatigue life was crucial for structural integrity assessment, yet the variability in material responses under complex loading paths made it challenging for both classical and data-driven models to achieve high accuracy. To address this issue, a contrastive learning-based framework was proposed in this study, enabling the construction of more generalized low-dimensional feature representations across different loading paths. This framework enhanced the robustness of fatigue life prediction without relying on mechanical assumptions. Experimental validation demonstrated that, compared to existing methods, the contrastive learning model learned more suitable feature encodings, significantly improving prediction performance. This framework provided a reference solution for engineering applications requiring reliability assessment under multiaxial stress conditions.

Keywords: Contrastive learning; Deep learning; Feature engineering; Life prediction; Multiaxial fatigue

***Corresponding author:**
Xingyue Sun
(xysun@nwpu.edu.cn)

Citation: Cui Z, Sun X, Chen X. Improvement of multiaxial fatigue life prediction performance based on contrastive learning feature extraction. *Int J AI Mater Design*. 2025;2(1):54-72.
doi: 10.36922/IJAMD025040004

Received: January 22, 2025

Revised: March 6, 2025

Accepted: March 14, 2025

Published online: March 28, 2025

1. Introduction

In modern high-tech industries such as electronic packaging, aerospace, and nuclear power generation, the role of multiaxial fatigue analysis has become increasingly critical.[1-5] Electronic packaging materials are subjected to complex thermal and mechanical loads, which can precipitate premature material and structural failures, thereby severely damaging structural integrity.[6,7] Consequently, comprehensive research into multiaxial fatigue is essential for enhancing the reliability and service life of electronic devices at the design and serving stages. In addition, precise assessments of fatigue behavior under complex stress environments are crucial for ensuring the safe operation of aerospace vehicles, nuclear power stations, and other fields.[8-10] Such studies not only facilitate a better understanding of material responses under multiaxial stresses but also advance the design of materials and the evaluation of structural integrity, pivotal for the development of safer and more efficient technological solutions.

Currently, with the development of advanced intelligent algorithms, data-driven methods have been widely developed and applied in the field of material fatigue life prediction.[11-18] From classical shallow machine learning algorithms to deep learning models based on neural networks, these methods have demonstrated excellent performance in various fatigue-related problems.[19-24] Jiang *et al.*[25] proposed a physics-informed multilayer nested neural network framework, using stacking fault energy, strain amplitude, strain rate, and temperature as input features. Physical constraints were embedded in the loss function to ensure the model adhered to known physical laws. The performance of the model was validated on fatigue data of 316 stainless steel. Zhu *et al.*[26] developed a deep learning model called Multi-GAT (Multi-Graph Attention Network) for predicting the high-cycle fatigue (HCF) life of titanium alloys. This model is integrated with an attention mechanism and uses a graph structure as the data structure, allowing the full consideration of relationships between samples. This approach enables the accurate prediction of HCF life for various titanium alloys using a limited number of sample data. Liao *et al.*[27] proposed a path-dependent adaptive physics-informed neural network to address the non-proportionality caused by phase differences. The model embedded multiple critical plane models into the loss function and achieved optimization through dynamic weight adjustment. Genetic algorithms and a meta-learning framework were used to optimize the weight hyperparameters. The meta-learning framework enabled the weights of different physical terms in the loss function to dynamically adapt based on the load path information. The explored meta-learning framework, applied through transfer learning to predict the fatigue life of 316L and 304 stainless steels, demonstrated strong generalization capabilities. Chen *et al.*[28] proposed a multiview neural network model incorporating frequency domain analysis. This model integrates convolutional neural networks (CNN), long short-term memory networks, and FNet with frequency domain analysis in a parallel structure, extracting effective features from the material loading path to predict fatigue life. Through ablation experiments, the extrapolation capability of the model was verified using specific test datasets. Zhang *et al.*[29] used a SHapley Additive exPlanations-informed recursive feature elimination method to identify key features in a multiaxial fatigue experiment dataset. Symbolic regression was employed to extract and encapsulate expressions predicting fatigue life based on these salient features, which were then integrated into the traditional mean squared error (MSE). This significantly improved the predictive accuracy of the model on the existing database.

In fatigue life prediction, constructing appropriate input features is crucial for improving the model's predictive performance.[30-33] Simply using raw experimental features as inputs may not lead to optimal results. To achieve effective feature engineering, researchers have drawn on various traditional empirical models, such as critical plane models,[34,35] damage mechanics models,[36,37] and fracture mechanics models.[38,39] These empirical insights help in selecting and designing more relevant features, which significantly enhance the model's ability to predict fatigue life accurately. Figure 1 provides an intuitive comparison between traditional physics-based models and data-driven approaches. In traditional methods, domain experts manually extract key features using physics-

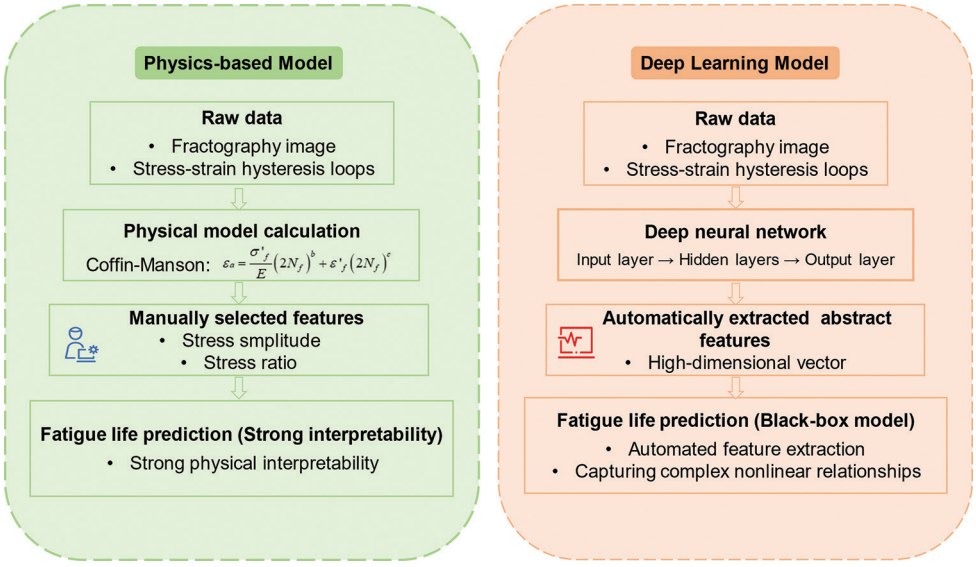

Figure 1. A comparison of fatigue life prediction frameworks: physics-based model versus deep learning model

based, whereas deep learning methods can automatically learn high-dimensional abstract representations from raw experimental data, capturing complex nonlinear relationships within the data. This comparison highlights the strengths and limitations of both approaches and underscores the importance of selecting appropriate input features for accurate fatigue life prediction. Wang *et al.*[40] proposed two physics-guided machine learning frameworks based on the Paris law and machine learning models. The first framework transforms the original features into new features using the Paris law, which are then combined with the original features and input into the machine learning model to predict fatigue life. The second framework integrates the Paris law and machine learning models using the Kalman filter, leveraging the advantages of both approaches and integrating information from different models. This enables the fusion of physical information and machine learning, allowing the model to account for factors ignored by physics-based models while ensuring consistency with physical results. Dong *et al.*[41] unified the influence of different defects through equivalent damage area representation using the M-integral fatigue model. By taking cyclic loading and equivalent damage area as inputs and fatigue life as output, the approach effectively improved the generalization ability and prediction accuracy of incomplete fatigue data. Fan *et al.*[42] expanded the original dataset using the *Z*-parameter model by restricting key parameters such as the size of critical defects, the relative depth of critical micro-defects, and stress levels within certain ranges. These extended datasets were then used to build machine learning models for ultra-high-cycle fatigue life prediction. With an appropriate increase in the size of the training set, the model demonstrated significantly improved accuracy and exhibited higher accuracy and stronger generalization compared to physical models. Gan *et al.*[43] used the Ye-Wang damage theory to derive theoretical estimates of material behavior, constructing additional features closely related to the desired outputs. This approach integrated original data information with domain knowledge. In addition, output standards were set to provide information for the data-driven process of model training and prediction, highlighting its potential in addressing small-sample problems. Wang *et al.* [44] characterized the fatigue life of additive manufacturing (AM) parts under different stress amplitudes using predictions from continuous damage mechanics associated with AM. These predictions, along with initial features such as experimental conditions, mechanical properties, porosity analysis, and surface morphology, were used as inputs. By learning the dependence on physical principles, the model can better map the nonlinear relationships between inputs and outputs.

The essence of feature engineering lies in the extraction and dimensionality reduction of complex, high-dimensional features, which brings samples closer together in high-dimensional space. Traditional models, based on physical laws and phenomenon analysis, focus on extracting and analyzing key values. However, many of these models are empirical, summarizing patterns that may not truly reflect the relationships between data samples, and they may not have clear physical models to describe the underlying dynamics. From the perspective of data relationships, dimensionality reduction or clustering techniques, such as K-means,[45] variational autoencoder,[46] and contrastive learning,[47-50] can also be employed with effective results. Among these, contrastive learning stands out by maximizing the consistency between similar samples and the disparity between different ones, enabling the learning of more robust and generalized feature representations. This approach does not require prior knowledge of label information, which reduces model usage costs. Furthermore, its adversarial sample pair learning strategy enhances the model's generalization ability. By optimizing the relative distances between samples, contrastive learning provides a powerful learning mechanism for complex datasets.[51] This is especially beneficial in unsupervised and self-supervised learning scenarios, where it outperforms traditional algorithms. These advantages have led to contrastive learning demonstrating outstanding application performance in a variety of fields, such as visual recognition,[52,53] natural language processing,[54,55] and sound analysis.[56,57]

This paper addresses the issue of multiaxial fatigue life prediction in materials by applying contrastive learning to effectively extract sample features across different multiaxial loading paths. The paper compares the feature representation performance of different model frameworks and tests various downstream task models. Compared to other clustering or dimensionality reduction algorithms, contrastive learning consistently achieves good prediction results, providing a new feature engineering strategy for multiaxial fatigue life prediction.

2. Data

2.1. Experimental data

In this study, 20 multiaxial fatigue experimental data of 316L stainless steel were used for training and validation of the contrastive learning model and downstream fatigue life prediction model. The experimental details can be found in the referenced published literature.[16,17] As shown in Figure 2, the experimental data included four typical uniaxial and multiaxial loading paths, each with five different amplitudes, with a stress ratio of $R = -1$. Table 1

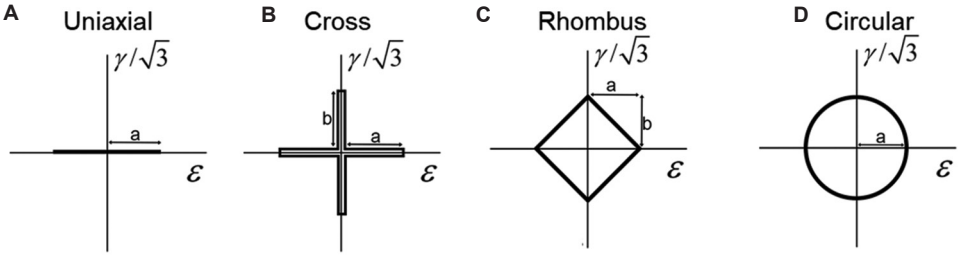

Figure 2. The diagram of multiaxial loading paths: (A) uniaxial, (B) cross, (C) rhombus, and (D) circular

Table 1. The detailed information of specimen for each loading path type

Loading path	ε_a	γ_a	N_f
Uniaxial	0.2	-	161000
	0.4	-	11865
	0.6	-	3339
	0.8	-	1719
	1.0	-	906
Cross	0.3	0.3	10023
	0.35	0.35	5831
	0.4	0.4	2038
	0.5	0.5	1378
	0.6	0.6	1311
Rhombus	0.3	0.3	7193
	0.35	0.35	4041
	0.4	0.4	3944
	0.5	0.5	1031
	0.6	0.6	795
Circular	0.3	0.3	5045
	0.35	0.35	2963
	0.4	0.4	1802
	0.5	0.5	744
	0.6	0.6	396

provides the amplitude and life information for each experimental specimen.

2.2. Data augmentation

To ensure the effectiveness of contrastive learning, data augmentation was also applied to enhance the training samples. The generative adversarial network (GAN)-based augmentation method used was derived from our previous research.[16] By integrating CNN with the Fourier transform, the proposed model achieves synergistic augmentation of stress-strain hysteresis loops and fatigue life prediction. Based on the data augmentation capability of this model, the 20 experimental samples were expanded to over 1000 samples for each loading path. For each multiaxial loading path, the strain amplitude of the augmented samples was distributed between the maximum and minimum values of the experimental data. Based on these augmented samples, 100 samples were uniformly selected for each loading path. A total of 400 augmented samples were selected for the training of the contrast learning model. Data augmentation effectively increases the training sample size and improves the performance of the contrastive learning model. In addition, only stress-strain hysteresis loops are required, without the need for fatigue life data, which reduces the difficulty of using augmented samples. In the subsequent downstream fatigue life prediction task, there are only the original 20 experimental datasets used for the training. In addition, in future work, constitutive material models or finite element simulations can also be employed as alternative methods for data augmentation.

2.3. Preprocessing

According to the previous introduction, the stress-strain hysteresis loop will serve as input features for the training of following contrastive learning and machine learning model. The output feature of the downstream supervised model is the logarithmic fatigue life. It can avoid the effect of its large magnitude range. Furthermore, all input and output features are normalized using the z-score method to ensure that the model captures the relative relationships between features rather than being influenced by their absolute values. The *z*-score normalization is described as Equation I:

$$\bar{x} = \frac{x - \mu_x}{\sigma_x} \quad \text{(I)}$$

Where $\bar{x}$ represents the normalized data, x represents the raw data, μ_x is the mean of the raw data, and σ_x represents the standard deviation of the entire sample space.

3. Algorithm

3.1. Contrastive learning architecture

Contrastive learning is a self-supervised learning strategy that enables the model to learn to distinguish between

similar and dissimilar samples to extract features, without relying on manually labeled data.[58] In the contrastive learning framework, a pair of samples (positive and negative pairs) is typically constructed, where positive pairs consist of similar or related samples and negative pairs consist of dissimilar samples. The model's objective is to minimize the distance between positive pairs and maximize the distance between negative pairs. This learning approach effectively utilizes unlabeled data and enhances the model's generalization ability on real-world data. Contrastive learning has been widely applied to various tasks, such as image recognition,[59,60] natural language processing,[61,62] and audio analysis,[63,64] demonstrating outstanding performance. Especially in the field of image processing, contrastive learning learns powerful visual feature representations by leveraging samples generated from different perspectives, sizes, or other transformations of images.

Among all contrastive learning models, SimCLR[65] learns representations by applying contrastive loss in the feature space to maximize the consistency between different augmented samples of the same data sample. The model framework, as shown in Figure 3, consists of four components: the data augmentation module, where each sample is randomly augmented to generate two related samples x_i and x_j, which are then passed through an encoder to extract high-dimensional features. After passing through a projection layer, the features are mapped to a lower-dimensional space, where the contrastive loss between the samples is computed to achieve maximum consistency.

In this work, contrastive learning was applied to the multiaxial fatigue life prediction of materials, using stress-strain hysteresis data as the input. Considering that contrastive learning can reduce the distance between similar samples in the feature space, this framework aims to extract general, discriminative features that are shared by samples with the same amplitude but different loading paths. These features were then used to construct a downstream fatigue life prediction model. The core idea of contrastive learning was implemented by constructing positive and negative sample pairs. The construction of these pairs was done by the data augmentation module, which randomly selected a small batch of data containing N samples. Negative samples were not specifically sampled, and after data augmentation, $2N$ data samples were generated. In this case, for each positive sample pair, there were $2(N-1)$ corresponding negative pairs. In this study, data augmentation involved adding Gaussian noise and applying time masking. Each sample was augmented with Gaussian noise, and with a 50% probability, time masking was applied, which randomly masked a certain number of time steps in the stress-strain data. After feature extraction through the encoder and projection layer, the distance between the features of positive and negative samples was computed. The loss function was minimized to learn a good feature space representation. Since SimCLR does not impose specific constraints on the network architecture, its performance improves after adding a projection layer. The model framework used in this study is shown in Figure 4, which adopts a synchronized symmetric architecture.

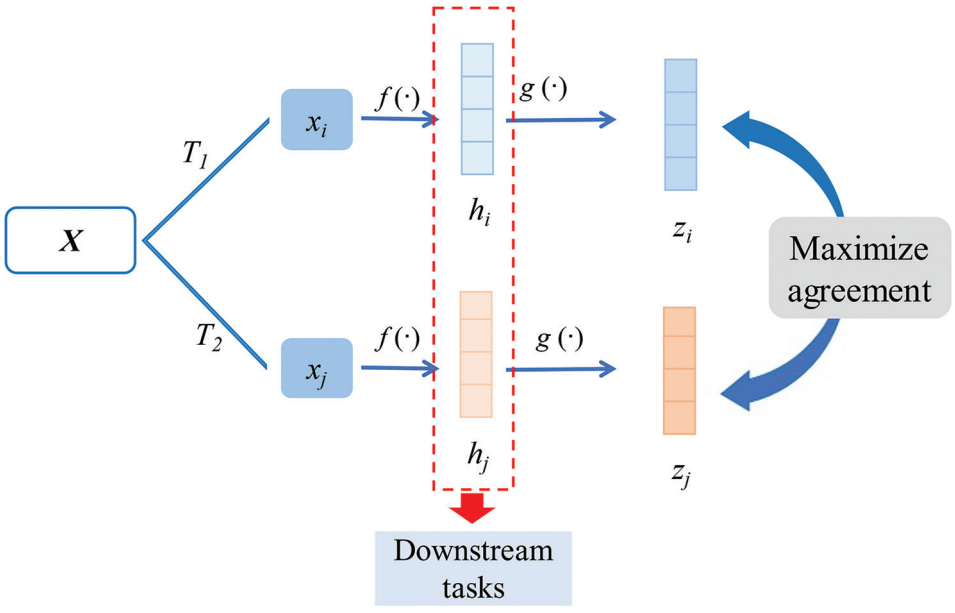

Figure 3. Schematic diagram of contrastive learning algorithm

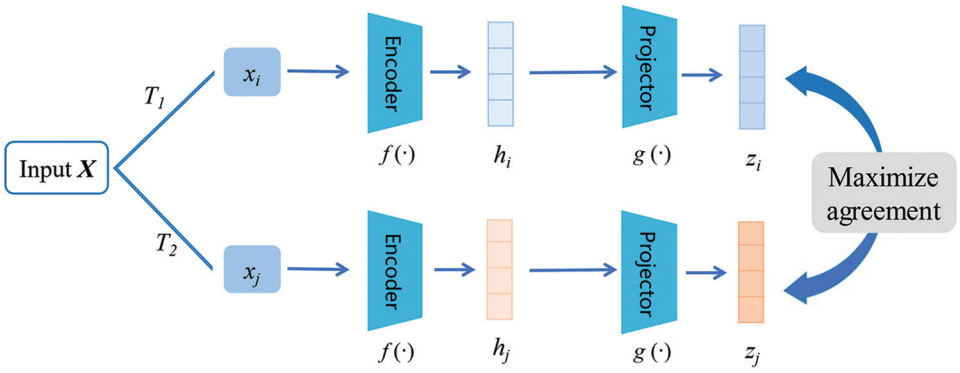

Figure 4. The framework of the proposed contrastive learning model

The core of contrastive learning is to optimize the model through a set of loss functions, enabling it to distinguish between similar and dissimilar samples. Below are some key formulas commonly used in contrastive learning. Among them, Equations II and IV are the formulas for the first sample generated by data augmentation, passed through the encoder and projection layer, respectively. Equation III and V are the formulas for the second sample x_j generated by data augmentation, after being processed by the encoder and projection layer. Equation VI is the formula for the calculation of the loss function. The Normalized Temperature-scaled Cross Entropy (NT-Xent) loss function enhances sample utilization and feature extraction quality during the learning process by effectively leveraging the relative information between samples. As a result, it has been widely applied in self-supervised learning tasks across various domains, including image processing,[66] text processing,[67] and more.T.

$$h_i = f\left(T_1\left(X\right);\Theta\right) = ReLU(\mathrm{W}_2^f\ f_1\left(\mathrm{T}_1\left(X\right);\ \Theta_1 + b_2^f\right) \quad \text{(II)}$$

$$h_i = f\left(T_2\left(X\right);\Theta\right) = ReLU(\mathrm{W}_2^f f_1\left(\mathrm{T}_2\left(X\right);\ \Theta_1 + b_2^f\right) \quad \text{(III)}$$

$$z_i = g\left(h_i\right) = W_2^g\ \sigma\left(W_1 h_i + b_1\right) + b_2 \quad \text{(IV)}$$

$$z_j = g\left(h_i\right) = W_2^g\ \sigma\left(W_1 h_j + b_1\right) + b_2 \quad \text{(V)}$$

$$l_{i,j} = -\log\frac{\exp(smi(z_i, z_j)/\tau)}{\sum_{k=1}^{2N} 1[k \neq 1]\exp(smi(z_i, z_j)/\tau)} \quad \text{(VI)}$$

where $\Theta = \{\Theta_1, \Theta_2\}$ represents the set of all learnable parameters of the encoder layers; W^f, W^g and b^f, b^g represent the weight matrices and biases of the encoder and projection layers, respectively; $1[k \neq 1]$ is an indicator function, which takes the value 1 when $k \neq i$; and τ represents the temperature parameter.

3.2. Downstream life prediction model

After contrastive learning, deep feature representations can be extracted from the material stress-strain hysteresis loop, and the similarity between samples from different loading paths can be enhanced. This greatly benefits subsequent multiaxial fatigue life prediction tasks. As shown in Figure 5, after obtaining these features, this study uses them as input for various supervised learning models, such as linear regression, support vector machines (SVM), eXtreme Gradient Boosting (XGBoost), or Artificial Neural Network (ANN), for fatigue life prediction. This approach not only improves the predictive accuracy of the model but also enhances its generalization ability when confronted with unseen complex loading conditions.

3.3. Evaluation criteria

In this study, the root MSE (RMSE) was used to describe the deviation between the predicted logarithmic fatigue life values and the experimental logarithmic fatigue life, as shown in Equation VII:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{i,pre} - y_{i,exp})^2} \quad \text{(VII)}$$

where $y_{i,\mathrm{pre}}$ represents the model-predicted logarithmic fatigue life value and $y_{i,\mathrm{exp}}$ represents the experimental logarithmic fatigue life value. In addition, the model's prediction performance was also evaluated from other aspects, including the distribution of the predicted fatigue life values and the experimental values.

4. Results and discussion

In this section, the proposed contrastive learning framework was evaluated by comparing different network architectures to explore models with feature representation capabilities. It also compares these models with other unsupervised learning clustering models in terms of

their ability to represent features effectively. Finally, based on the learned feature representations, a performance comparison is conducted for downstream life prediction tasks, examining the predictive effectiveness across various machine learning models.

4.1. Performance with different encoder network architectures

In this section, the experiment first determined two hyperparameters of the proposed contrastive learning framework: the number of layers in the encoder and the output feature dimension. The framework of the contrastive learning model is shown in Figure 6, where the encoder was set to two layers, and the output dimension was uniformly set to 128, which preliminarily validates the model's effectiveness. From a time-series perspective, the stress-strain hysteresis data recorded the change of stress and strain over time during one complete cycle. Therefore, to further determine which encoder structure was more suitable for the original data, four structures—One-Dimensional Convolutional Neural Network (1D-CNN), Two-Dimensional Convolutional Neural Network (2D-CNN), Gated Recurrent Unit (GRU), and ANN—were used. The high-dimensional features extracted by the encoder were passed through two fully connected layers and uniformly reduced to 64 dimensions, and training was conducted on the stress-strain hysteresis data. Table 2 provides the detailed hyperparameters of the contrastive learning models with different network frameworks.

In this study, to ensure the effectiveness of model training, the original dataset (with 20 samples in total) was randomly split into training and testing sets at a 6:4 ratio. Although the original samples had an equal number of samples for each type, random sampling without stratified control led to some distribution bias between different

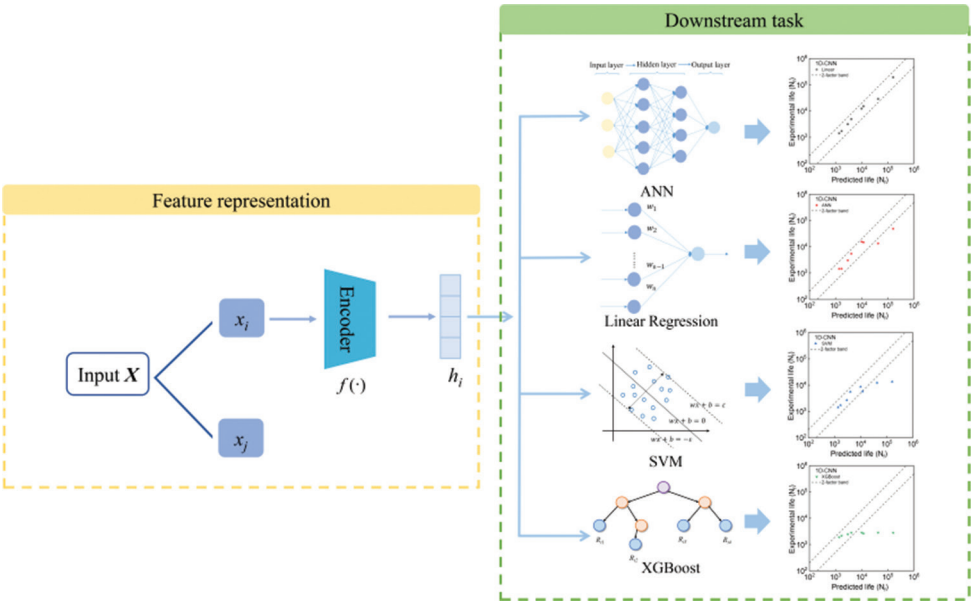

Figure 5. The utilization of deep features provided by contrastive learning model for downstream fatigue life prediction
Abbreviations: ANN: Artificial neural network; SVM: Support vector machine; XGBoost: eXtreme gradient boosting.

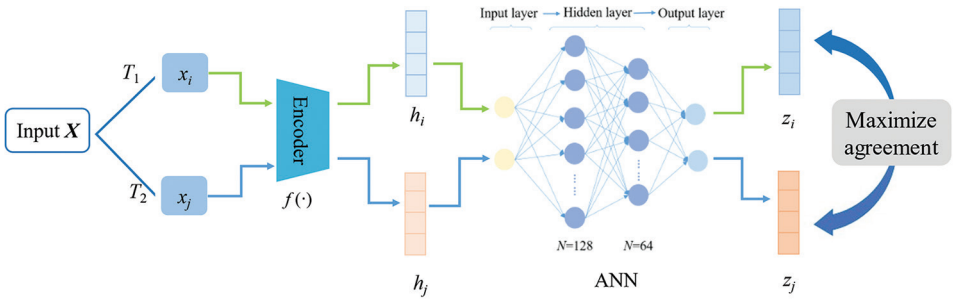

Figure 6. Hyperparameters and schematic diagrams of contrastive learning model with different network architectures
Abbreviation: ANN: Artificial neural network.

types of data in the training and testing sets, simulating the uneven distribution of data in real-world conditions. To further mitigate the impact of this distribution imbalance, additional sample data generated by a GAN model was included in the training set to augment the samples and improve the model's generalization ability. Each sample underwent two different data augmentation methods. The training loss of each contrastive learning model framework is shown in Figure 7. After 150 epochs, all models converged well, validating the effectiveness of the proposed framework.

Table 2. The detailed hyperparameters of contrastive learning models with different network architectures

Encoder	Projector	Hyperparameters			
		Configuration	Stride	Padding	τ
1D-CNN	ANN	Convolutional Layers: 2, Kernel Size: (3,3), Filters: 64, 128	1	1	0.3
2D-CNN		Convolutional Layers: 2, Kernel Size: (3, 3), Filters: 64, 128	1	1	
GRU		GRU Units: 128 Number of layers: 2	\	\	
ANN		Neurons: 256, 128	\	\	

Abbreviations: 1D-CNN: One-Dimensional Convolutional Neural Network; 2D-CNN: Two-Dimensional Convolutional Neural Network; ANN: Artificial Neural Network; GRU: Gated Recurrent Unit.

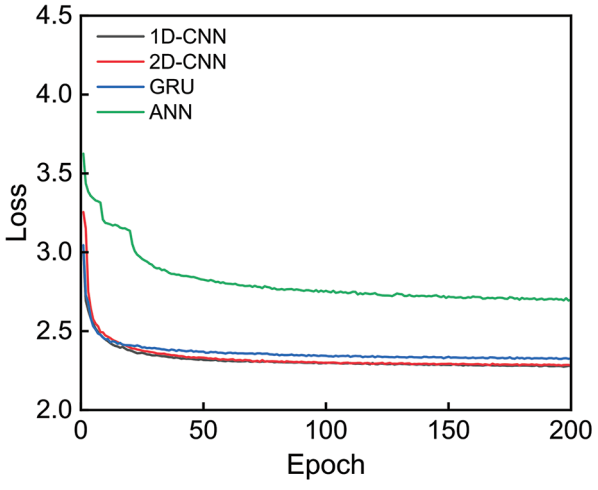

Figure 7. The training process of contrastive learning model between different architecture models

Abbreviations: 1D-CNN: One-dimensional convolutional neural network; 2D-CNN: Two-dimensional convolutional neural network; ANN: Artificial neural network; GRU: Gated recurrent unit.

To investigate whether the model successfully learned the distribution of data features during training, a visual analysis of the input and output features before and after training was performed. Specifically, in the visualization process, original data was used as the visualization samples, excluding the augmented samples generated by GAN. The t-distributed stochastic neighbor embedding (t-SNE)[68] method was employed for dimensionality reduction and visualization of the features, with the pre- and post-training visualization results shown in Figures 8 and 9. In these figures, data points with the same color represented features of samples with the same amplitude but different loading paths, while different colors corresponded to samples with different amplitudes. The visualization results indicated that in the original data, samples from the same loading path were clustered together in the feature space. However, for features generated by each encoder from the initial samples, the distribution of features for the same amplitude data was quite disordered, and no obvious clustering or separation of samples was observed in the reduced-dimensional space. It was worth noting that, although there was no clear separation of feature types in the visualization space, the learned features still showed significant application effects in downstream tasks. This may have been because t-SNE's low-dimensional visualization lost important information from the high-dimensional space, and the visualization results can not accurately reflect the intrinsic structure of the high-dimensional feature space.

4.2. The effectiveness of downstream fatigue life prediction

To further find the optimal downstream regression model, the linear regression model was considered, as it is simple,

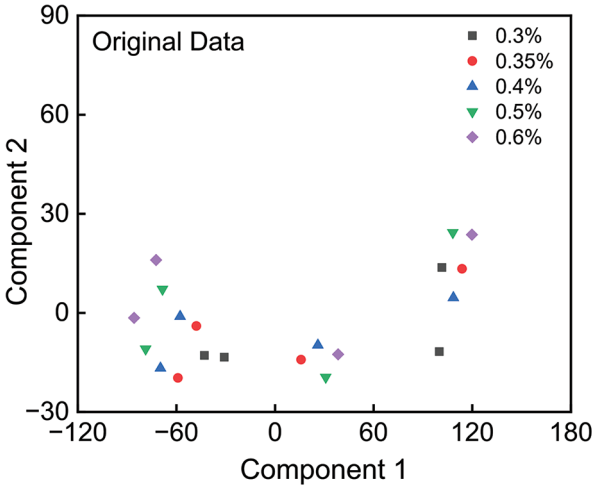

Figure 8. Visualization results of the original data

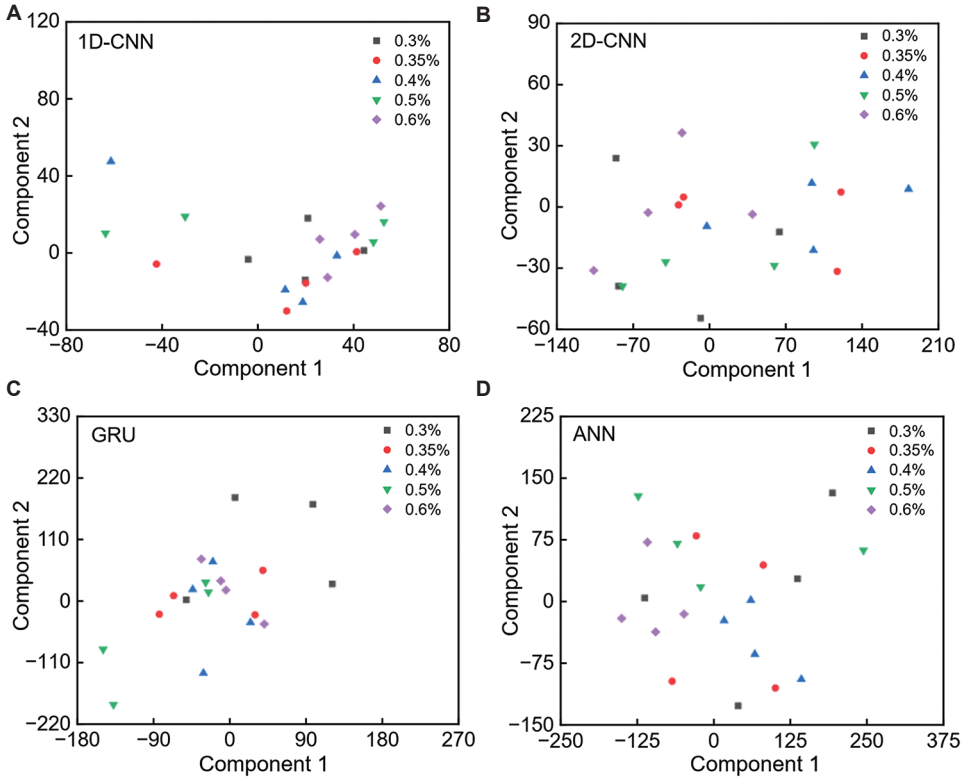

Figure 9. Visualization results of the feature representations obtained through contrastive learning with different encoder architectures: (A) 1D-CNN, (B) 2D-CNN, (C) GRU, and (D) ANN.

Abbreviations: 1D-CNN: One-dimensional convolutional neural network; 2D-CNN: Two-dimensional convolutional neural network; ANN: Artificial neural network; GRU: Gated recurrent unit.

efficient to train, and suitable as a baseline model to provide a reference for more complex models. In addition, since the downstream training set only contained a small amount of data randomly split 60:40 from the original dataset, and given the robust performance of SVM and XGBoost in small sample scenarios, as well as their capability in handling nonlinear problems, three models—linear regression, SVM, and XGBoost—were chosen for training. As a widely used machine learning model, the ANN model was also selected as one of the models for downstream tasks. The features learned by contrastive learning from different network architectures were used as input for the downstream models, and the models were trained. The RMSE of the experimental results on the test set is shown in Figure 10. In the figure, the *x*-axis represents the contrastive learning models with different network architectures, and each bar color represents a different downstream model. The *y*-axis represents the RMSE value, with lower RMSE indicating better prediction performance. The results showed that, for all contrastive learning models regardless of the framework, the extracted features had relatively lower RMSE values on the linear regression model in the downstream task. For all downstream linear regression models, whether using 1D-CNN or 2D-CNN frameworks in the contrastive learning model, the extracted features achieved lower RMSE values. In particular, the contrastive learning model with 1D-CNN as the encoder achieved the lowest RMSE value on the downstream task. This suggests that the features extracted by 1D-CNN are sufficiently simple, have high linear separability, and can be effectively utilized by linear models.

In addition, the predicted fatigue life on the test set and the experimental fatigue life are shown in Figure 11. It can be observed that the best performance was achieved when 1D-CNN was used as the encoder. In contrast to 1D-CNN, the contrastive learning model with 2D-CNN as the encoder had one test point lying outside the 2-factor band of the linear regression model. However, for the XGBoost model, the features extracted by contrastive learning performed the worst on the downstream model, with the RMSE significantly higher than that of other models. This might be because the XGBoost model was more suited to handle high-dimensional complex features, and it was unable to fully leverage the advantages of the features learned by contrastive learning. In addition, during computation, it might have introduced extra noise or information loss, severely impacting fatigue life prediction performance.

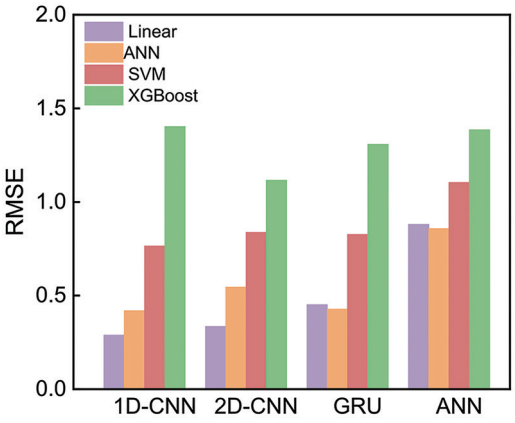

Figure 10. The RMSE performances of life prediction models. Abbreviations: 1D-CNN: One-dimensional convolutional neural network; 2D-CNN: Two-dimensional convolutional neural network; ANN: Artificial neural network; GRU: Gated recurrent unit; Linear: Linear regression; RMSE: Root mean squared error; SVM: Support vector machine; XGBoost: eXtreme Gradient boosting.

Based on these results, CNN1D was selected as the encoder for the contrastive learning framework, and the linear regression layer was used as the downstream regression model for fatigue life prediction. This combination not only achieved the lowest RMSE value but also provided acceptable prediction results on the test set, fully validating the superiority of this combination.

To further validate the effectiveness of data augmentation in contrastive learning on the training results, the CNN1D-based contrastive learning encoder was chosen, and the performance of the downstream model was compared under two conditions: with and without data augmentation. The training loss is shown in Figure 12. It can be observed that with data augmentation, the model achieved a sufficiently small loss after fewer epochs and began to converge quickly. In contrast, the model without data augmentation had a relatively large initial loss and required more epochs to converge.

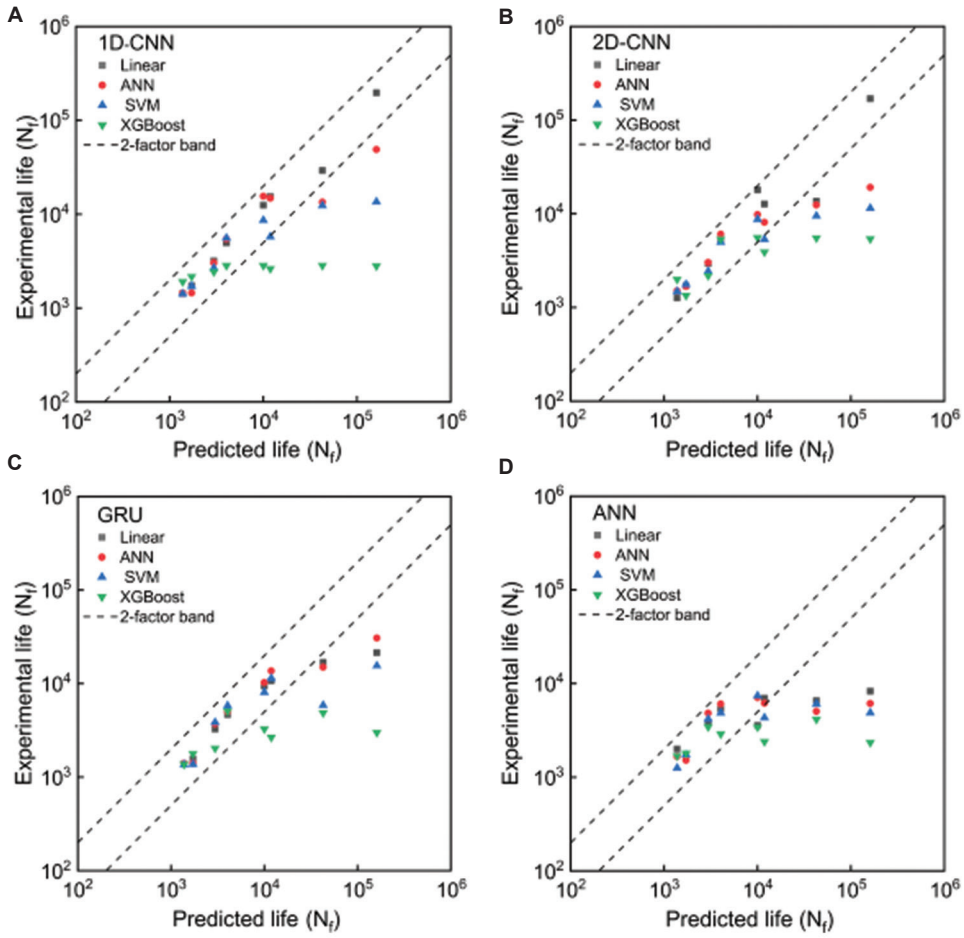

Figure 11. The detailed predicted results of contrastive learning and downstream models: (A) 1D-CNN, (B) 2D-CNN, (C) GRU, and (D) ANN. Abbreviations: 1D-CNN: One-dimensional convolutional neural network; 2D-CNN: Two-dimensional convolutional neural network; ANN: Artificial neural network; GRU: Gated recurrent unit; Linear: Linear regression; SVM: Support vector machine; XGBoost: eXtreme gradient boosting.

The predicted fatigue life on the test set and the comparison with experimental results, as well as the RMSE, are shown in Figures 13 and 14. From the figures, the results indicated that while the model without data augmentation had only two data points outside the 2-factor band, the model with data augmentation produced better prediction results and had a lower RMSE. Specifically, the RMSE for the model with data augmentation was reduced by 16.35% compared to the model without data augmentation.

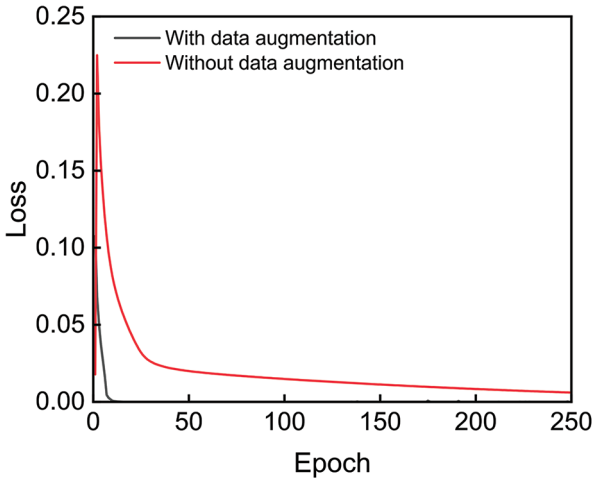

Figure 12. The evolution of loss function during training loss of contrastive learning models with and without data augmentation

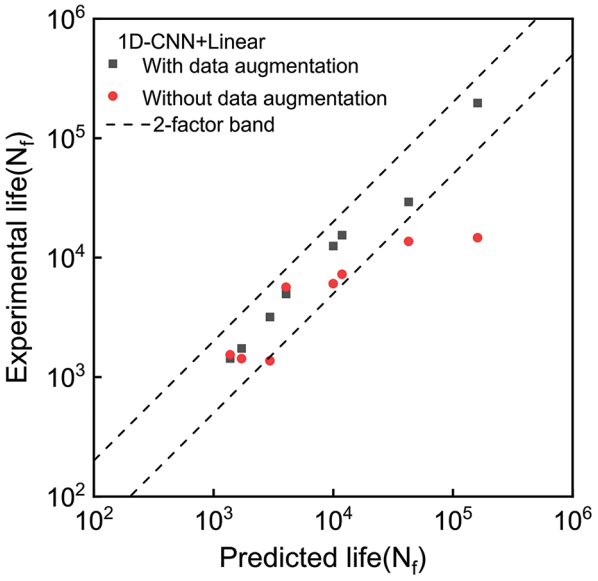

Figure 13. The predicted results of 1D-CNN contrastive learning encoder and linear regression downstream model with and without data augmentation.
Abbreviations: 1D-CNN: One-dimensional convolutional neural network; Linear: Linear regression.

In addition, to investigate the effectiveness of the contrastive learning framework, this experiment explored the scenario where no contrastive learning was used, and the stress-strain data was directly input into the downstream regression model, with data augmentation applied. The data augmentation mainly involved using GAN-generated data to expand the dataset, which, especially when the sample size was small, can increase the diversity and generalization ability of the data, thereby enhancing the model's predictive performance. The choice of downstream model was consistent with previous experiments, and the model's performance was compared using RMSE and prediction results, as shown in Figure 15.

From the experimental results, it was be observed that when contrastive learning was not used but data augmentation was applied, the performance of the four models was poor. The minimum RMSE value was close to 0.7, and in the comparison of the model's predictions with experimental results, the best performance still had four data points outside the 2-factor band. Whether in terms of RMSE or the comparison between predicted and experimental values, the model performance was worse than that of the model using deep features extracted by the contrastive learning framework. Especially for the linear regression model, it showed the best performance when contrastive learning was applied, while its performance was the worst when contrastive learning was not used.

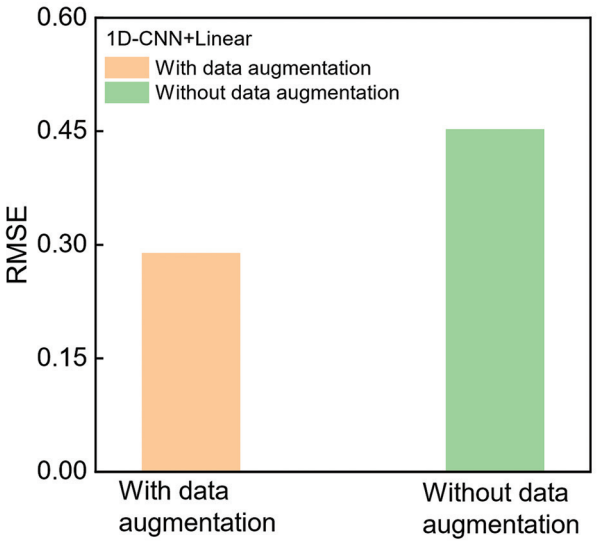

Figure 14. The RMSE performances of 1D-CNN contrastive learning encoder and linear regression downstream model with and without data augmentation.
Abbreviations: 1D-CNN: One-dimensional convolutional neural network; Linear: Linear regression; RMSE: Root mean squared error.

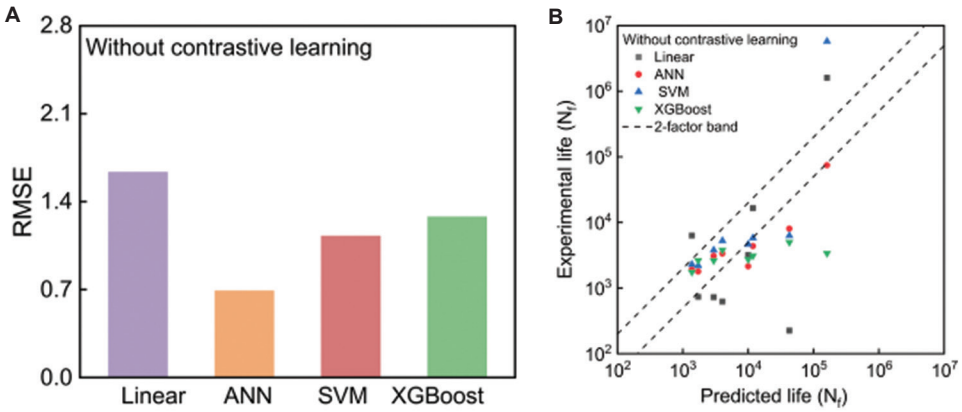

Figure 15. The performance of the downstream model in the case of data augmentation without contrastive learning. (A) The RMSE of the downstream models. (B) The prediction results of the downstream model.

Abbreviations: ANN: Artificial neural network; Linear: Linear regression; RMSE: Root mean squared error; SVM: Support vector machine; XGBoost: eXtreme gradient boosting.

This indicated that although data augmentation can provide more training data, the augmented data might only be generated based on surface-level features. The four downstream models were unable to solely rely on augmented data to understand the underlying structure and complex patterns in the data. In contrast, the contrastive learning model, by maximizing the similarity between similar samples and maximizing the distance between dissimilar samples, continuously optimized the representation space of the data. The large amount of data provided by augmentation can help the model learn in a broader sample space, enabling the model to learn more universal and representative features. Therefore, although pure data augmentation did not provide sufficient structural information, the training strategy of contrastive learning allowed the model to better uncover the inherent relationships within the data, thereby improving the performance of the downstream model.

To further verify the superiority of the combination of contrastive learning framework and data augmentation, the performance of the downstream regression model was investigated in this experiment without the contrastive learning framework and data augmentation. The input to the downstream model did not undergo any form of data augmentation, nor was it trained with features learned through the contrastive learning framework. Instead, the raw stress-strain data was used as input, and the logarithmic fatigue life was used as the output. As a result, the model's performance was directly constrained by the data volume and sample diversity. The model's RMSE and prediction results were compared with the experimental results, as shown in Figure 16.

From the figure, it was evident that the RMSE of the models was above 0.5, and in the comparison between predicted and experimental fatigue life, although Linear Regression models only had one point outside the 2-factor band, most of the data within the error margin was distributed near its edges. This indicated that while the models could, to some extent, capture some patterns in the data through simple training methods, the performance remained limited, and the internal relationships were not fully exploited. In contrast, under the contrastive learning framework, the linear regression model had the smallest RMSE and its predicted values were well-distributed along the diagonal, outperforming all other downstream models.

When compared with the second experiment, despite the absence of the contrastive learning framework, the model still relied on the limited information from the raw data for training. The model depended on the data's quality and complexity to learn some effective features. This proved that simply relying on data augmentation did not necessarily contribute positively to model performance. Although data augmentation could increase the training sample size, the augmented data did not add meaningful information. Without an effective training strategy, the augmented data could introduce significant noise and negatively affect the model's performance, diminishing the effectiveness of data augmentation.

Through these three experiments, the effects of data augmentation, contrastive learning framework, and their combination on downstream models were explored. The experimental results not only showed the effects of each factor individually but also demonstrated the synergistic effect when they were combined. Ultimately, the experiments confirmed the superiority of the combination of contrastive learning framework and data augmentation. While data augmentation could effectively increase the sample size, it might introduce noise and did not necessarily contribute positively to the model's training process, and

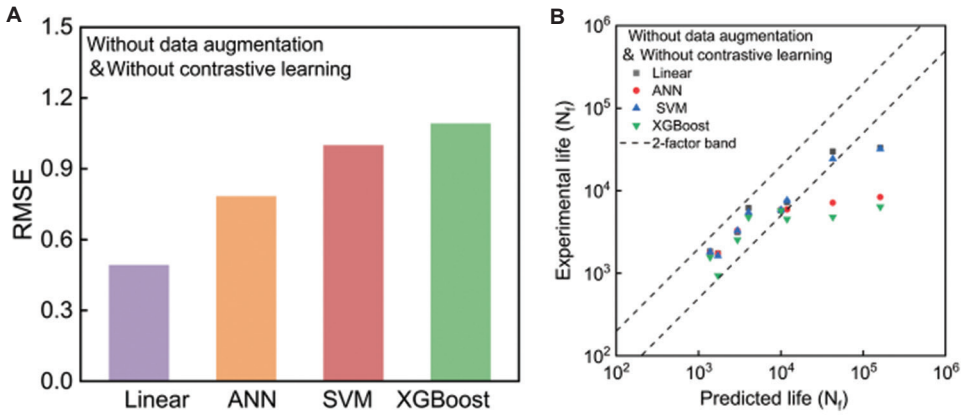

Figure 16. The performance of the downstream models without data augmentation without contrastive learning. (A) The RMSE of the models. (B) The prediction results of the models.
Abbreviations: ANN: Artificial neural network; Linear: Linear regression; RMSE: Root mean squared error; SVM: Support vector machine; XGBoost: eXtreme gradient boosting.

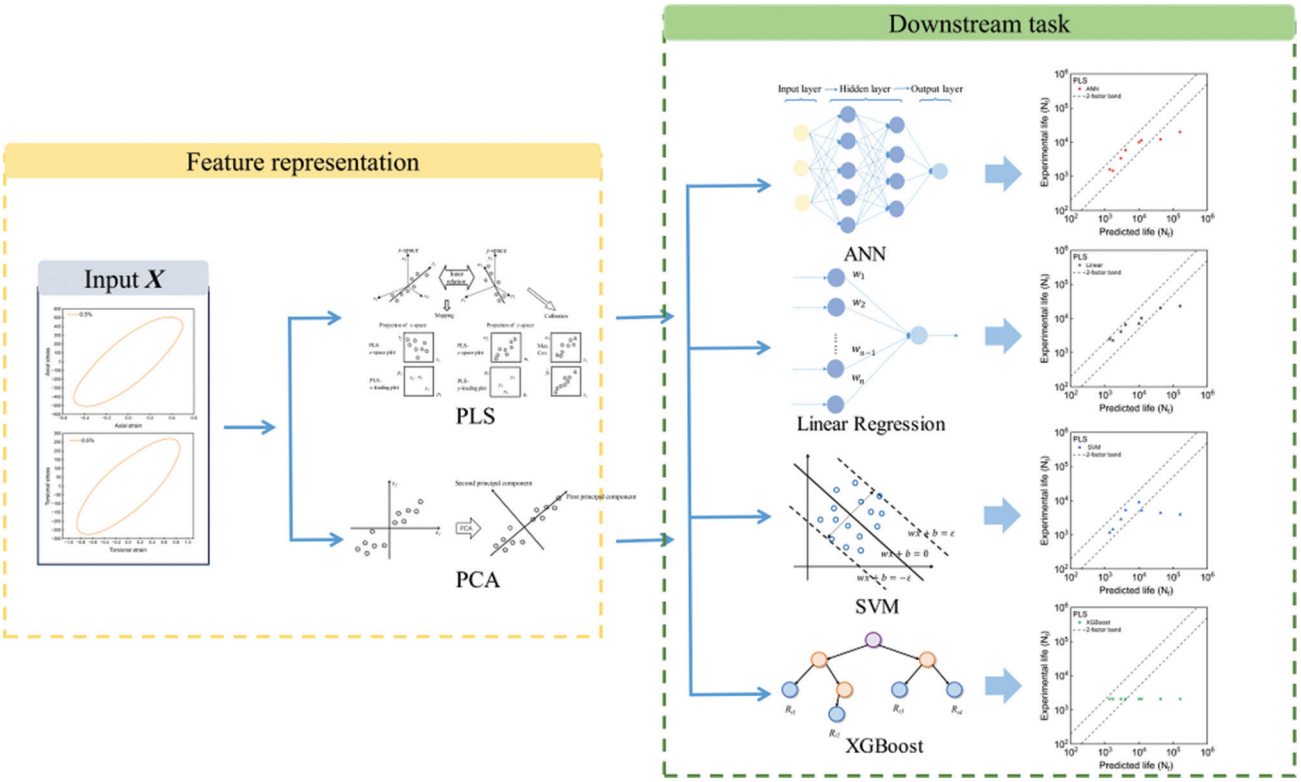

Figure 17. Schematic diagrams of classical unsupervised clustering learning algorithms
Abbreviations: ANN: Artificial neural network; PCA: Principal component analysis; PLS: Partial least squares; SVM: Support vector machine; XGBoost: eXtreme gradient boosting.

might even interfere with the model's training. In contrast, the combination of data augmentation and the contrastive learning framework's training strategy maximized the utilization of data samples, extracted deep features from the stress-strain data, and applied them to the downstream model training, achieving the best prediction results.

4.3. Comparison of the effectiveness between different clustering methods

In this section, two unsupervised learning algorithms, partial least squares (PLS) and principal component analysis (PCA), were applied to perform dimensionality reduction on the

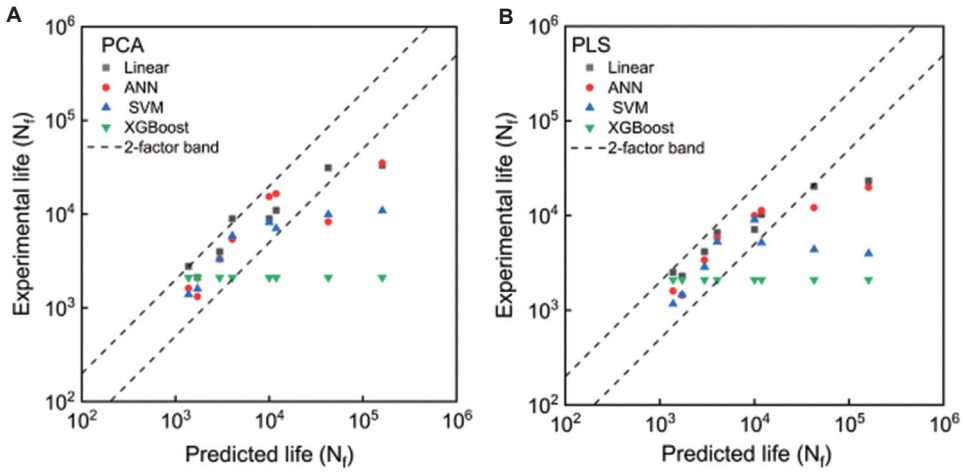

Figure 18. Downstream prediction results of other unsupervised algorithms: (A) PCA, and (B) PLS
Abbreviations: ANN: Artificial neural network; Linear: Linear regression; PCA: Principal component analysis; PLS: Partial least squares; SVM: Support vector machine; XGBoost: eXtreme gradient boosting.

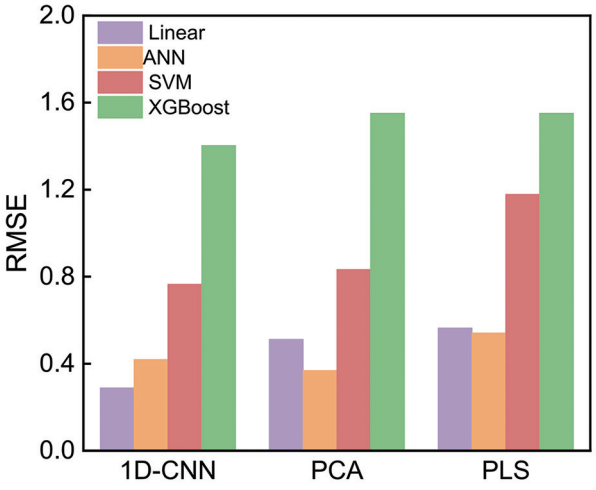

Figure 19. Comparison of downstream RMSE between other unsupervised learning algorithms and contrastive learning
Abbreviations: 1D-CNN: One-dimensional convolutional neural network; ANN: Artificial neural network; Linear: Linear regression; RMSE: Root mean squared error; SVM: Support vector machine; XGBoost: eXtreme gradient boosting.

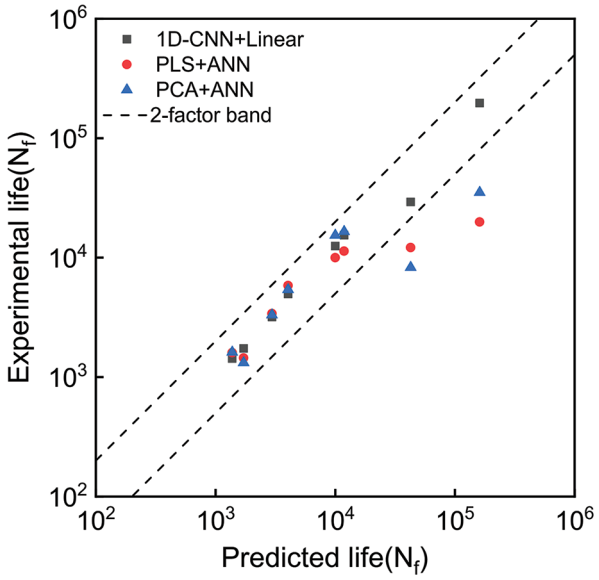

Figure 20. Comparison with the optimal performance prediction results of the contrastive learning framework
Abbreviations: 1D-CNN: One-dimensional convolutional neural network; ANN: Artificial neural network; Linear: Linear regression; PCA: Principal component analysis; PLS: Partial least squares.

original data. The reduced features were then fed into four regression models for training. The overall process of the unsupervised learning algorithms is illustrated in Figure 17. Using these two methods, the original data was reduced to a certain dimension, and the extracted features were directly applied to the construction of regression models. ANN, linear regression, SVM, and XGBoost were still used as the regression models. The comparison between the predicted results from the test set and the experimental values is shown in Figure 18, and the RMSE is shown in Figure 19. From the results, it can be seen that the features reduced by both PLS and PCA performed relatively well on linear models and ANN, with two points falling outside the 2-factor band. However, even the most optimal PCA and PLS frameworks still did not outperform the performance of contrastive learning, where the contrastive learning framework remained the best, as shown in Figure 20. This indicated that, compared to the features extracted by unsupervised learning algorithms, the contrastive learning framework can improve prediction performance to some extent.

Table 3. The detailed number of each specimen and defects for each type

Feature extraction method	RMSE of the downstream prediction model			
	Linear	ANN	SVM	XGBoost
1D-CNN	**0.2889**	0.41946	0.76503	1.40253
2D-CNN	**0.33526**	0.54549	0.8378	1.11583
GRU	0.45183	**0.42774**	0.82675	1.30814
ANN	0.88021	**0.85771**	1.10393	1.3852
PLS	0.56382	**0.54084**	1.17852	1.55079
PCA	0.51162	**0.36833**	0.83287	1.55079

Note: The values in boldface represent the lowest RMSE among the four downstream models under the same conditions.
Abbreviations: 1D-CNN: One-dimensional convolutional neural network; 2D-CNN: Two-dimensional convolutional neural network; ANN: Artificial neural network; GRU: Gated recurrent unit; Linear: Linear regression; PCA: Principal component analysis; PLS: Partial least squares; RMSE: Root mean squared error; SVM: Support Vector Machines.

Table 4. RMSE results of ablation experiments on 1D-CNN with contrastive learning

Ablation setting	RMSE of the downstream prediction model			
	Linear	ANN	SVM	XGBoost
Without data augmentation	**0.45243**	-	-	-
Without contrastive learning	1.63522	**0.69097**	1.12755	1.28143
Without data augmentation & without contrastive learning	**0.49226**	0.78342	0.50171	1.09144
With data augmentation & with contrastive learning	**0.2889**	0.41946	0.76503	1.40253

Note: The values in boldface represent the lowest RMSE among the four downstream models under the same conditions.
Abbreviations: 1D-CNN: One-dimensional convolutional neural network; ANN: Artificial neural network; Linear: Linear regression; RMSE: Root mean squared error; SVM: Support vector machines; XGBoost: eXtreme gradient boosting.

Table 3 summarizes the RMSE values of different feature extraction methods across various downstream prediction models. It was observed that the features extracted by the contrastive learning framework with 1D-CNN as the encoder achieved the best performance on the downstream linear regression model. To further validate the contributions of data augmentation and contrastive learning, ablation experiments were conducted. Table 4 presents the results of removing data augmentation, removing contrastive learning, and removing both, illustrating their impact on model performance.

5. Conclusion

In this study, a multiaxial fatigue hysteresis feature extraction method based on contrastive learning is proposed. This method effectively extracts deep feature representations from complex multiaxial fatigue stress-strain responses, which are then utilized in downstream fatigue life prediction tasks to enhance prediction accuracy. The specific conclusions are as follows:

(i) Compared to network architectures such as ANN, GRU, and 2D-CNN, the 1D-CNN network achieves the best contrastive learning performance and is more stable during training. The deep feature representations extracted through contrastive learning, when visualized by t-SNE, show a chaotic distribution in the reduced-dimensional space, with no obvious clustering or separation of categories. However, the extracted features enable the downstream fatigue life prediction model to more easily learn from samples with different loading paths, achieving excellent performance even with a simple linear regression model.

(ii) Compared to other unsupervised learning algorithms, the features extracted using contrastive learning show better similarity and consistency. Through comparative experiments, contrastive learning is found to be more effective in extracting features related to fatigue life prediction from stress-strain hysteresis data, thereby helping downstream models better uncover the underlying patterns in the data. Compared to traditional unsupervised learning algorithms, contrastive learning demonstrates stronger robustness and effectiveness when handling multiaxial fatigue data.

(iii) The feature representations learned through contrastive learning exhibit superior predictive performance in downstream tasks. In multiple machine learning models, contrastive learning consistently achieves better prediction results. Compared to scenarios without contrastive learning, the maximum reduction in RMSE in models such as SVM, ANN, and others can reach 86.26%. In addition, the prediction stability is improved, as evidenced by a reduction in the standard deviation of repeated experiments.

(iv) Contrastive learning has the potential to be further extended for applications in multiaxial fatigue life prediction and similar domains. Leveraging the benefits of contrastive learning, it can help achieve few-shot or even zero-shot learning for downstream tasks. This approach can also contribute to addressing challenges in fields such as electronic packaging and multiscale structural integrity, where data scarcity and the need for robust predictive models are key concerns.

(v) To further enhance the effectiveness and applicability of the proposed framework, several key directions warrant exploration. One important area is integrating contrastive learning with traditional physics-based models to bridge data-driven insights with mechanical

principles, thereby improving model interpretability. In addition, extending this framework to materials with heterogeneous microstructures can help evaluate its performance under different fatigue mechanisms and further validate its generalization capability.

Acknowledgments

None.

Funding

The authors gratefully acknowledge financial support for this work from the National Natural Science Foundation of China (No. 12302098).

Conflict of interest

The authors declared that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Author contributions

Conceptualization: Xingyue Sun
Formal analysis: Ziyu Cui
Investigation: Xingyue Sun, Ziyu Cui
Methodology: Xingyue Sun
Writing – original draft: Ziyu Cui
Writing – review & editing: Xingyue Sun, Xu Chen

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

The data are available from the corresponding author upon reasonable request.

References

1. Towashiraporn P, Gall K, Subbarayan G, *et al*. Power cycling thermal fatigue of Sn-Pb solder joints on a chip scale package. *Int J Fatigue*. 2024;26(5):497-510.

   doi: 10.1016/j.ijfatigue.2003.09.004

2. Donnerbauer K, Bill T, Starke P, *et al*. Fatigue life evaluation of metastable austenitic stainless steel AISI347 based on nondestructive testing methods for different environmental conditions. *Int J Fatigue*. 2024;179:108056.

   doi: 10.1016/j.ijfatigue.2023.108056

3. Zhao B, Xie L, Song J, *et al*. Fatigue life prediction of aero-engine compressor disk based on a new stress field intensity approach. *Int J Mech Sci*. 2020;165:105190.

   doi: 10.1016/j.ijmecsci.2019.105190

4. Liang Q, Peng C, Li X. A multi-state Semi-Markov model for nuclear power plants piping systems subject to fatigue damage and random shocks under dynamic environments. *Int J Fatigue*. 2023;168:107448.

   doi: 10.1016/j.ijfatigue.2022.107448

5. Yu Z, Sun X, Xing R, Chen X. Unified prediction of uniaxial ratcheting deformation at elevated temperatures with physics-informed multimodal network. *Int J Plast*. 2025;187:104275.

   doi: 10.1016/j.ijplas.2025.104275

6. Maniar Y, Konstantin G, Sharma A, *et al*. Solder joint lifetime modeling under random vibrational load collectives. *JOM*. 2020;72(2):898-905.

   doi: 10.1007/s11837-019-03947-1

7. Wijker JJ. *Spacecraft Structures*. Berlin: Springer Science & Business Media; 2008.

8. Min KD, Lee BS, Kim SJ. Effects of oxide on fatigue crack growth behaviour of type 347 stainless steel in PWR water conditions. *Fatigue Fract Eng Mater Struct*. 2015;38(8):960-969.

   doi: 10.1111/ffe.12290

9. Agency IAE. *Assessment and Management of Ageing of Major Nuclear Power Plant Components Important to Safety: PWR Pressure*, IAEA-TECDOC-1556, IAEA, Vienna; 2007.

10. Guo C, Yu D, Sun X, *et al*. Fatigue failure mechanism and life prediction of a cast duplex stainless steel after thermal aging. *Int J Fatigue*. 2021;146:106161.

    doi: 10.1016/j.ijfatigue.2021.106161

11. Kishore P, Mondal A, Trivedi A, *et al*. A microstructure sensitive machine learning-based approach for predicting fatigue life of additively manufactured parts. *Int J Fatigue*. 2025;192:108724.

    doi: 10.1016/j.ijfatigue.2024.108724

12. Yang J, Kang G, Liu Y, Kan Q. A novel method of multiaxial fatigue life prediction based on deep learning. *Int J Fatigue*. 2021;151:106356.

    doi: 10.1016/j.ijfatigue.2021.106356

13. Zhang M, Sun CN, Zhang X, *et al*. High cycle fatigue life prediction of laser additive manufactured stainless steel: A machine learning approach. *Int J Fatigue*. 2019;128:105194.

    doi: 10.1016/j.ijfatigue.2019.105194

14. Li J, Yang Z, Qian G, Berto F. Machine learning based very-high-cycle fatigue life prediction of Ti-6Al-4V alloy fabricated by selective laser melting. *Int J Fatigue*. 2022;158:106764.

    doi: 10.1016/j.ijfatigue.2022.106764

15. Jing G, Ma T, Wang Z, *et al*. Physical hierarchical neural

network for low cycle fatigue life prediction of compacted graphite cast iron based on small data. *Int J Fatigue*. 2024;188:108509.

doi: 10.1016/j.ijfatigue.2024.108509

16. Sun X, Zhou K, Shi S, Song K, Chen X. A new cyclical generative adversarial network based data augmentation method for multiaxial fatigue life prediction. *Int J Fatigue*. 2022;162:106996.

doi: 10.1016/j.ijfatigue.2022.106996

17. Sun X, Zhou T, Song K, Chen X. An image recognition based multiaxial low-cycle fatigue life prediction method with CNN model. *Int J Fatigue*. 2023;167:107324.

doi: 10.1016/j.ijfatigue.2022.107324

18. Zhou T, Sun X, Yu Z, Chen X. A generalization ability-enhanced image recognition based multiaxial fatigue life prediction method for complex loading conditions. *Eng Fract Mech*. 2024;295:109802.

doi: 10.1016/j.engfracmech.2023.109802

19. Wang H, Zhang J, Li B, *et al*. Machine learning-based fatigue life prediction of laser powder bed fusion additively manufactured Hastelloy X via nondestructively detected defects. *Int J Struct Integr*. 2024;16(1):104-126.

doi: 10.1108/IJSI-09-2024-0161

20. Zhao F, Cui J, Yuan M, Zhao J. A weakly supervised pairwise comparison learning approach for bearing health quantitative evaluation and remaining useful life prediction. *Eng Computat*. 2023;40(7/8):1593-1616.

doi: 10.1108/EC-12-2022-0747

21. Mao M, Wang W, Lu C, Jia F, Long X. Machine learning for board-level drop response of BGA packaging structure. *Microelectron Reliabil*. 2022;134:114553.

doi: 10.1016/j.microrel.2022.114553

22. Long X, Lu C, Shen Z, Su Y. Identification of Mechanical Properties of Thin-Film Elastoplastic Materials by Machine Learning. *Acta Mechanica Solida Sinica*. 2023;36(1):13-21.

doi: 10.1007/s10338-022-00340-5

23. Long X, Mao M, Lu C, Li C, Li R. Modeling of heterogeneous materials at high strain rates with machine learning algorithms trained by finite element simulations. *J Micromech Mol Phys*. 2021;6(1):2150001.

doi: 10.1142/S2424913021500016

24. Cao W, Sun X, Li Y, *et al*. Multiaxial damage parameter evaluation by neural network-based symbolic regression. *Eng Fract Mechan*. 2025;315:110809.

doi: 10.1016/j.engfracmech.2025.110809

25. Jiang L, Hu Y, Liu Y, *et al*. Physics-informed machine learning for low-cycle fatigue life prediction of 316 stainless steels. *Int J Fatigue*. 2024;182:108187.

doi: 10.1016/j.ijfatigue.2024.108187

26. Zhu S, Zhang Y, Zhu B, *et al*. High cycle fatigue life prediction of titanium alloys based on a novel deep learning approach. *Int J Fatigue*. 2024;182:108206.

doi: 10.1016/j.ijfatigue.2024.108206

27. Liao H, Pan J, Su X, Sun X, Chen X. A path-dependent adaptive physics-informed neural network for multiaxial fatigue life prediction. *Int J Fatigue*. 2025;193:108799.

doi: 10.1016/j.ijfatigue.2024.108799

28. Chen S, Zhou X, Bai Y. A frequency domain enhanced multi-view neural network approach to multiaxial fatigue life prediction for various metal materials. *Int J Fatigue*. 2025;190:108620.

doi: 10.1016/j.ijfatigue.2024.108620

29. Zhang P, Tang K, Wang A, Wu H, Zhong Z. Neural network integrated with symbolic regression for multiaxial fatigue life prediction. *Int J Fatigue*. 2024;188:108535.

doi: 10.1016/j.ijfatigue.2024.108535

30. Long X, Li H, Iyela PM, Kang SB. Predicting the bond stress-slip behavior of steel reinforcement in concrete under static and dynamic loadings by finite element, deep learning and analytical methods. *Eng Fail Anal*. 2024;161:108312.

doi: 10.1016/j.engfailanal.2024.108312

31. Zhou K, Sun X, Shi S, Song K, Chen X. Machine learning-based genetic feature identification and fatigue life prediction. *Fatigue Fract Eng Mater Struct*. 2021;44(9):2524-2537.

doi: 10.1111/ffe.13532

32. Zhou T, Sun X, Chen X. A multiaxial low-cycle fatigue prediction method under irregular loading by ANN model with knowledge-based features. *Int J Fatigue*. 2023;176:107868.

doi: 10.1016/j.ijfatigue.2023.107868

33. Long X, Ding X, Li J, Dong R, Su Y, Chang C. Indentation reverse algorithm of mechanical response for elastoplastic coatings based on LSTM deep learning. *Materials*. 2023;16(7):2617.

doi: 10.3390/ma16072617

34. Wang CH, Brown MW. A path-independent parameter for fatigue under proportional and non-proportional loading. *Fatigue Fract Eng Mater Struct*. 1993;16(12):1285-1297.

doi: 10.1111/j.1460-2695.1993.tb00739.x

35. Fatemi A, Socie DF. A critical plane approach to multiaxial fatigue damage including out-of-phase loading. *Fatigue Fract Eng Mater Struct*. 1988;11(3):149-165.

doi: 10.1111/j.1460-2695.1988.tb01169.x

36. Zhan Z, Hu W, Li B, Zhang Y, Meng Q, Guan Z. Continuum damage mechanics combined with the extended finite element method for the total life prediction of a metallic

component. *Int J Mech Sci*. 2017;124-125:48-58.

doi: 10.1016/j.ijmecsci.2017.03.002

37. Yang S, Hu W, Meng Q, Zhao B. A new continuum damage mechanics-based two-scale model for high-cycle fatigue life prediction considering the two-segment characteristic in S-N curves. *Fatigue Fract Eng Mater Struct.* 2020;43(2):387-402.

doi: 10.1111/ffe.13161

38. Brown MW, Miller KJ. A theory for fatigue failure under multiaxial stress-strain conditions. *Proc Inst Mech Eng.* 1973;187(1):745-755.

doi: 10.1243/pime_proc_1973_187_161_02

39. Carraro PA, Quaresimin M. A damage based model for crack initiation in unidirectional composites under multiaxial cyclic loading. *Compos Sci Technol*. 2014;99:154-163.

doi: 10.1016/j.compscitech.2014.05.012

40. Wang L, Zhu S, Luo C. Physics-guided machine learning frameworks for fatigue life prediction of AM materials. *Int J Fatigue*. 2023;172:107658.

doi: 10.1016/j.ijfatigue.2023.107658

41. Dong Y, Yang X, Chang D, Li Q. Predicting fatigue life of multi-defect materials using the fracture mechanics-based physics-informed neural network framework. *Int J Fatigue*. 2025;190:108626.

doi: 10.1016/j.ijfatigue.2024.108626

42. Fan JL, Zhu G, Zhu ML, Xuan FZ. A data-physics integrated approach to life prediction in very high cycle fatigue regime. *Int J Fatigue*. 2023;176:107917.

doi: 10.1016/j.ijfatigue.2023.107917

43. Gan L, Wu H, Zhong Z. On the use of data-driven machine learning for remaining life estimation of metallic materials based on Ye-Wang damage theory. *Int J Fatigue*. 2022;156:106666.

doi: 10.1016/j.ijfatigue.2021.106666

44. Wang H, Li B, Xuan FZ. Fatigue-life prediction of additively manufactured metals by continuous damage mechanics (CDM)-informed machine learning with sensitive features. *Int J Fatigue*. 2022;164:107147.

doi: 10.1016/j.ijfatigue.2022.107147

45. Cam LML, Neyman J. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. United States: University of California Press; 1967.

46. Kingma DP, Welling M. Auto-Encoding Variational Bayes. *In Proceedings of the 2nd International Conference on Learning Representations (ICLR2014)*; 2014.

doi: 10.48550/arXiv.1312.6114

47. Zhang CS, Chen J, Li QL, Deng BQ, Wang J, Chen CC. Deep contrastive learning: A survey. *Acta Automatica Sin*. 2023;49(1):15-39.

doi: 10.16383/j.aas.c220421

48. Jaiswal A, Babu AR, Zadeh MZ, Banerjee D, Makedon F. A survey on contrastive self-supervised learning. *Technologies*. 2021;9(1):2.

doi: 10.3390/technologies9010002

49. Le-Khac PH, Healy G, Smeaton AF. Contrastive representation learning: A framework and review. *IEEE Access*. 2020;8:193907-193934.

doi: 10.1109/ACCESS.2020.3031549

50. Liu X, Zhang F, Hou Z, *et al*. Self-supervised Learning: Generative or Contrastive. In: *IEEE Transactions on Knowledge and Data Engineering*; 2021. p. 1-1.

doi: 10.1109/TKDE.2021.3090866

51. van den Oord A, Li Y, Vinyals O. *Representation Learning with Contrastive Predictive Coding*; 2019. *arXiv*. Available from: https://arxiv.org/abs/1807.03748 [Last accessed on 2025 Jan 08].

52. He K, Fan H, Wu Y, *et al*. Momentum Contrast for Unsupervised Visual Representation Learning. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; 2020. p. 9726-9735. Available from: https://ieeexplore.ieee.org/document/9157636 [Last accessed on 2025 Jan 08].

53. Grill JB, Strub F, Altché F, *et al*. Bootstrap your own latent a new approach to self-supervised learning. In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc.; 2020. p. 21271-21284.

54. Radford A, Kim JW, Hallacy C, *et al*. Learning Transferable Visual Models from Natural Language Supervision. In: *International Conference on Machine Learning*. PMLR; 2021. p. 8748-8763.

55. Yang C, An Z, Zhou H, *et al*. Online knowledge distillation via mutual contrastive learning for visual recognition. *IEEE Trans Pattern Anal Mach Intell*. 2023;45(8):10212-10227.

doi: 10.1109/TPAMI.2023.3257878

56. Baevski A, Zhou H, Mohamed A, Auli M. wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations. *Proceedings of the 34th International Conference on Neural Information Processing Systems (NeurIPS 2020)*; 2020. p. 12449-12460.

57. Hershey JR, Chen Z, Le Roux J, Watanabe S. Deep clustering: Discriminative embeddings for segmentation and separation. In: *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; 2016. p. 31-35. Available from: https://ieeexplore.ieee.org/abstract/document/7471631 [Last accessed on 2025 Jan 13].

58. Hu H, Wang X, Zhang Y, Chen Q, Guan Q. A comprehensive survey on contrastive learning. *Neurocomputing*. 2024;610:128645.

doi: 10.1016/j.neucom.2024.128645

59. Gutmann M, Hyvärinen A. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. *JMLR Workshop and Conference Proceedings*; 2010. p. 297-304.

60. Caron M, Misra I, Mairal J, Goyal P, Bojanowski P, Joulin A. Unsupervised learning of visual features by contrasting cluster assignments. In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc., 2020. p. 9912-9924.

61. Reimers N, Gurevych I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics; 2019. p. 3982-3992.

62. Rusak E, Reizinger P, Juhos A, Bringmann O, Zimmermann RS, Brendel W. *InfoNCE: Identifying the Gap between Theory and Practice. arXiv*; 2024. Available from: https://arxiv.org/abs/2407.00143 [Last accessed on 2025 Jan 08].

63. Luo Y, Chen Z, Hershey JR, Le Roux J, Mesgarani N. Deep clustering and conventional networks for music separation: Stronger together. In: *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; 2017. p. 61-65. Available from: https://ieeexplore.ieee.org/abstract/document/7952118 [Last accessed on 2025 Jan 08].

64. Wu S, Wang Y, Jiang Y, Zhang Q, Liu J. CRATI: Contrastive representation-based multimodal sound event localization and detection. *Knowl Based Syst*. 2024;305:112692.

doi: 10.1016/j.knosys.2024.112692

65. Chen T, Kornblith S, Norouzi M, Hinton G. A simple framework for contrastive learning of visual representations. In: *Proceedings of the 37th International Conference on Machine Learning*. Vol. 119. JMLR; 2020.

66. Guo Q, Wang C, Xiao D, Huang Q. A lightweight open-world pest image classifier using ResNet8-based matching network and NT-Xent loss function. *Expert Syst Appl*. 2024;237:121395.

doi: 10.1016/j.eswa.2023.121395

67. Kim T, Yoo KM, Lee SG. Self-Guided Contrastive Learning for BERT Sentence Representations. In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*; 2021. p. 2528-2540.

doi: 10.18653/v1/2021.acl-long.197

68. van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res*. 2008;9(86):2579-2605.