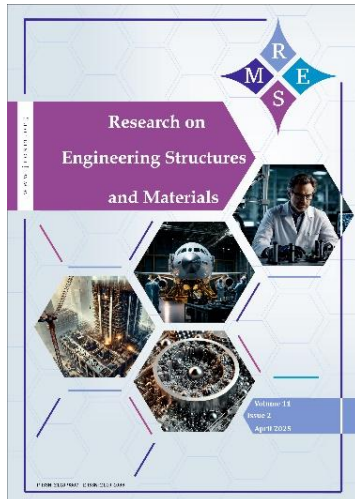




Intelligent prediction model of durability of shoe material foam structure integrating transformer attention mechanism and microscopic representation

Jinfeng Huang, Hangxi Fu

Online Publication Date: 20 September 2026

URL: <http://www.jresm.org/archive/resm2026-2065im0728rs.html>

DOI: <http://dx.doi.org/10.17515/resm2026-2065im0728rs>

Journal Abbreviation: *Res. Eng. Struct. Mater.*

To cite this article

Huang J, Fu H. Intelligent prediction model of durability of shoe material foam structure integrating transformer attention mechanism and microscopic representation. *Res. Eng. Struct. Mater.*, 2026; 12(5): 4283-4298.

Disclaimer

All the opinions and statements expressed in the papers are on the responsibility of author(s) and are not to be regarded as those of the journal of Research on Engineering Structures and Materials (RESM) organization or related parties. The publishers make no warranty, explicit or implied, or make any representation with respect to the contents of any article will be complete or accurate or up to date. The accuracy of any instructions, equations, or other information should be independently verified. The publisher and related parties shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with use of the information given in the journal or related means.



Published articles are freely available to users under the terms of Creative Commons Attribution - NonCommercial 4.0 International Public License, as currently displayed at [here](https://creativecommons.org/licenses/by-nc/4.0/) (the "CC BY - NC").

Intelligent prediction model of durability of shoe material foam structure integrating transformer attention mechanism and microscopic representation

Jinfeng Huang ^{*,1,a}, Hangxi Fu ^{2,b}

¹Department of Arts and Crafts, Meizhouwan Vocational Technology College, Putian, 351119, China

²Department of Intelligent Manufacturing Engineering, Meizhouwan Vocational Technology College, Putian, 351119, China

Article Info

Article History:

Received 28 Jul 2026

Accepted 13 Sep 2026

Keywords:

Microscopic representation;
Attention mechanism;
Transformer;
Foaming structure;
Durability

Abstract

The durability evaluation of modern shoe material foaming media has long relied on physical testing and faces technical bottlenecks such as long cycles and difficulty in revealing microscopic failure mechanisms. In response to this problem, the research aims to explore an evaluation path that takes into account local morphological quantification and long-range force dependence analysis and constructs an intelligent durability prediction model that integrates the window transformer attention mechanism and micro topological dimensionality reduction representation. The results show that in the application of anti-fatigue topology optimization, the model successfully suppressed the median stress of the foam matrix from 8.7MPa to 3.4MPa and strictly controlled the local extreme value at 6.1 MPa, reducing simulated stress concentration. Facing a severely distorted scene with a porosity of 70%, the mean absolute error was stably suppressed at 5.1k cycles, and network-only inference required 32.6 ms, demonstrating excellent fault tolerance and computational economy. In addition, in the long-term dynamic variable load verification, the predicted deformation rate at the end of life (9.80%) was highly consistent with the FEA reference level (9.90%), showing close agreement with the simulated loading trajectory. The research has opened up the cross-scale entity mapping from porous physical network to macroscopic mechanics and established an intelligent digital template with great transformation value for high-throughput virtual concurrent screening of shoe material formulas and anti-fatigue reverse auxiliary engineering.

© 2026 MIM Research Group. All rights reserved.

1. Introduction

1.1 Motivation for The Proposed Research

With the evolution of modern shoe engineering, foamed Porous Media (PM) has been widely used due to its lightweight and high resilience properties. However, cell geometry and cyclic loading conditions influence the fatigue response of cellular structures [1]. Machine-learning-assisted formulation design can reduce experimental effort in developing polyurethane foams with targeted density and mechanical strength [2]. In recent years, deep learning methods based on conventional Convolutional Neural Networks (CNNs) have gradually been introduced into this intersection field. However, predicting the response of heterogeneous materials requires linking microstructural representations to macroscopic mechanical behavior [3]. Therefore, there is an urgent need to explore an intelligent regression mechanism that combines efficient microscopic feature extraction and global spatial dependence modeling capabilities to break through the existing technical bottleneck of durability assessment.

*Corresponding author: huang.jfeng@outlook.com

^aorcid.org/0009-0005-4845-5236; ^borcid.org/0009-0007-7369-9819

DOI: <http://dx.doi.org/10.17515/resm2026-2065im0728rs>

Res. Eng. Struct. Mat. Vol. 12 Iss. 5 (2026) 4283-4298

1.2 Related Work

To break through the above bottlenecks, existing research focuses on PM microscopic characterization and mechanical property prediction, exploring dimensions such as image analysis, feature extraction, and intelligent regression model construction, and has accumulated rich theoretical results. In terms of microscopic characterization, Nguyen and others reviewed image processing techniques for characterizing PM and examining pore morphology, liquid transport, and mechanical damage [4]. Furthermore, Zhuang et al. introduced deep learning and used CNNs to construct a bidirectional prediction framework of PM features and compression mechanical response, which significantly improved the evaluation accuracy [5]. In addition, Kadulkar et al. reviewed the use of machine learning in dimensionality reduction and accelerated material design [6]. However, when existing visual networks deal with complex pores, they are often limited by local receptive fields, making it difficult to capture long-range spatial dependencies such as stress concentration on pore walls. In response to this limitation, the study proposes an intelligent regression model that integrates Transformer Attention Mechanism (TAM) and Microscopic Characterization (MC) to break through the gap between global topology modeling and durability prediction.

As a cutting-edge architecture, TAM has been widely used in many fields such as vision, especially in MC. It has significant advantages in resolving long-range dependencies through self-attention, thus showing strong modeling potential. Khan et al. reviewed vision transformers and CNN-transformer hybrids, highlighting their approaches to visual representation learning [7]. Subsequently, in the field of microscopic image analysis, Alrfou and others reviewed deep learning methods, including CNNs and Transformers, for extracting microscopic features and modeling structure–property relationships [8]. Furthermore, in response to the pain points of limited local receptive fields of classic networks, Qin's team proposed a factorization architecture to reconstruct the TAM paradigm, which successfully captured long-range dependencies and enhanced robustness while maintaining low computational overhead [9]. In view of this, introducing TAM across fields into an intelligent prediction model for the durability of shoe material foam structures is highly feasible for fitting the global topological characteristics of complex pores.

MC has significant advantages in solving structure-activity mapping problems through quantitative analysis of topological structures. It has been widely used in materials science and other fields, especially in PM research, and has important value in macroscopic performance prediction. Sahoo and Mandal constructed three-dimensional carbon-foam models from observed microstructures and used finite element analysis to evaluate effective mechanical properties [10]. In porous materials, the Andrä H team combined three-dimensional images to simulate fiber networks, revealing the decisive influence of pore characteristics on macroscopic thermal conductivity [11]. Furthermore, Prifling et al. combined three-dimensional pore-space descriptors with regression models to predict gas diffusion in loam and sand [12]. Based on the above research logic, using MC to accurately extract morphological prior information is the only way to drive the intelligent prediction model of shoe material foam structure durability to approach the true mechanical limit.

1.3 Research Objective and Highlights

In summary, substantial progress has been made in PM morphology quantification and basic performance mapping. However, traditional CNNs are limited by their local receptive fields and have difficulty capturing the long-range dependence of pore wall stress concentration, resulting in insufficient model generalization prediction accuracy in complex scenarios. Therefore, the study proposes an Intelligent Prediction Model for Durability of Footwear Foam Structures (IPM-FMD) that combines TAM and MC. This model uses CNNs to extract morphological features from orthogonal distance-map channels and couples the resulting spatial feature maps with a self-attention network for fatigue-life regression. The contribution lies in preserving the spatial arrangement of distance-encoded pore morphology throughout CNN feature extraction and window-attention processing, followed by channel–spatial reweighting for fatigue-life regression. The role of the Swin Transformer (ST) is to model interactions among image features, while multi-dimensional attention adjusts their contributions to the regression output. Component

contributions are examined through the existing ablation variants, and engineering applications are evaluated through numerical simulations.

2. Methods and Materials

2.1 Overall Architecture of IPM-FMD Model

Efficient representations of heterogeneous microstructures should retain morphological information relevant to their effective mechanical properties [13]. To this end, a new morphological dimensionality reduction representation strategy is constructed through orthogonal cross-section sampling and coupled distance transformation. Its complete intelligent regression prediction link architecture is shown in Figure 1.

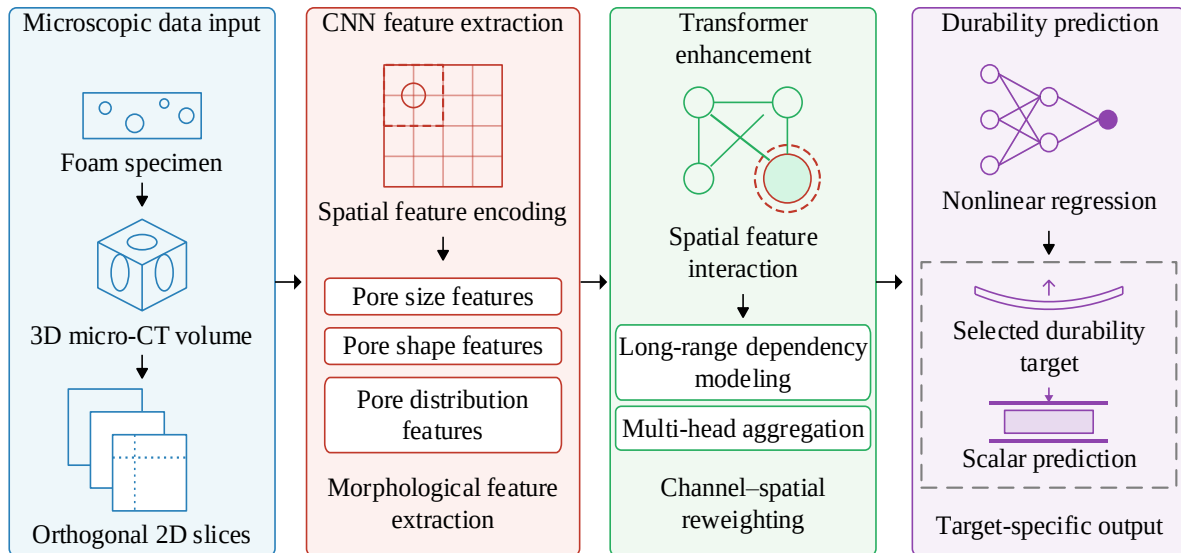


Fig. 1. Overall architecture diagram of IPM-FMD model

Note: One scalar is predicted for the selected durability target.

In Figure 1, IPM-FMD sequentially extracts three-dimensional orthogonal slices, uses CNNs to extract spatial morphological features, and relies on Swin Transformer blocks and multi-dimensional attention to model spatial interactions and reweight the learned features. Finally, the nonlinear regression network outputs a scalar prediction for the selected durability target.

2.2 Dimensionality Reduction Characterization of Microscopic Characteristics of Shoe Foam Pores Based on CNNs

2.2.1 Two-Dimensional Dimensionality Reduction and Pore Morphology Reconstruction of Shoe Material Foam Images

The fatigue performance of cellular lattice materials depends on unit-cell geometry and local stress concentration [14-15]. To reduce the computational cost of three-dimensional processing while retaining selected microscopic morphological priors, the model maps three-dimensional foaming pores to two-dimensional pseudo-color features, as shown in Figure 2.

In Figure 2, the model first receives a reconstructed three-dimensional tomographic volume, and achieves preliminary dimensionality reduction from three-dimensional voxels to a two-dimensional plane by extracting orthogonal slices through the internal pore space. For sliced images, intensity normalization and local contrast enhancement precede Li's minimum cross-entropy thresholding to separate the pore and solid phases. Subsequently, the Euclidean distance transformation matrix of this binary segmentation map is calculated to encode the distance from solid-phase pixels to the nearest segmented pore pixels. This representation describes local distance-to-pore information within the sampled planes. The distance field is calculated as shown in Equation (1).

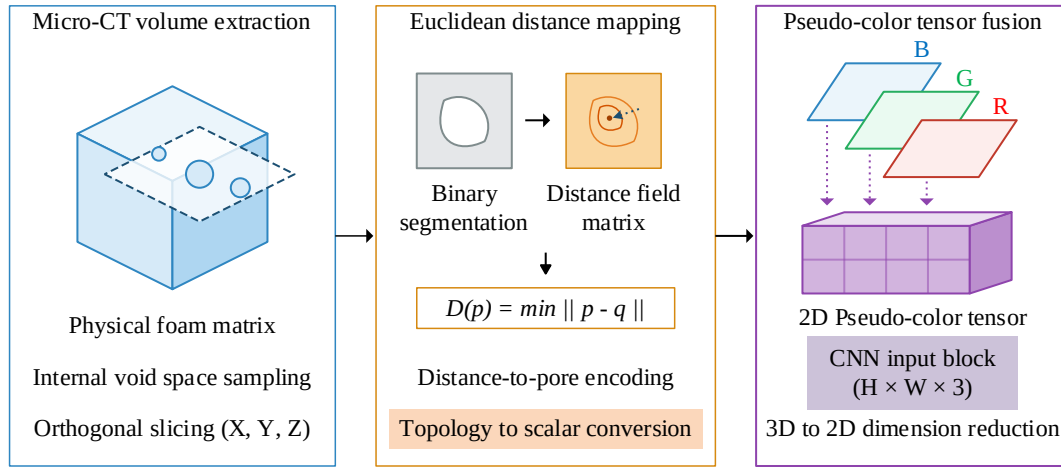


Fig. 2. Dimensionality reduction mapping from three-dimensional foaming pores to two-dimensional pseudo-color features

$$D_{map}(u, v) = \min_{(x, y) \in \Omega_{void}} \sqrt{(u - x)^2 + (v - y)^2} \quad (1)$$

In Equation (1), $D_{map}(u, v)$ is the Euclidean distance field value from the matrix pixel point with coordinate (u, v) on the orthogonal slice to the nearest pore boundary. Ω_{void} represents the microscopic pore physical space set obtained after segmentation by cross entropy threshold. (x, y) is the target pixel coordinate belonging to the pore set. Finally, the distance field matrix of multi-directional slices is channel-joined and reconstructed into a two-dimensional tensor in pseudo-color format, which is used as an efficient input feature for subsequent CNNs.

2.2.2 CNNs Convolution Extraction for Microscopic Pore Topological Features

Mechanical characterization of polymeric foams is complicated by their cellular morphology and deformation behavior [16]. To accurately capture these key anti-deformation features, the model introduces CNNs to perform deep down-sampling extraction. The specific network architecture is shown in Figure 3.

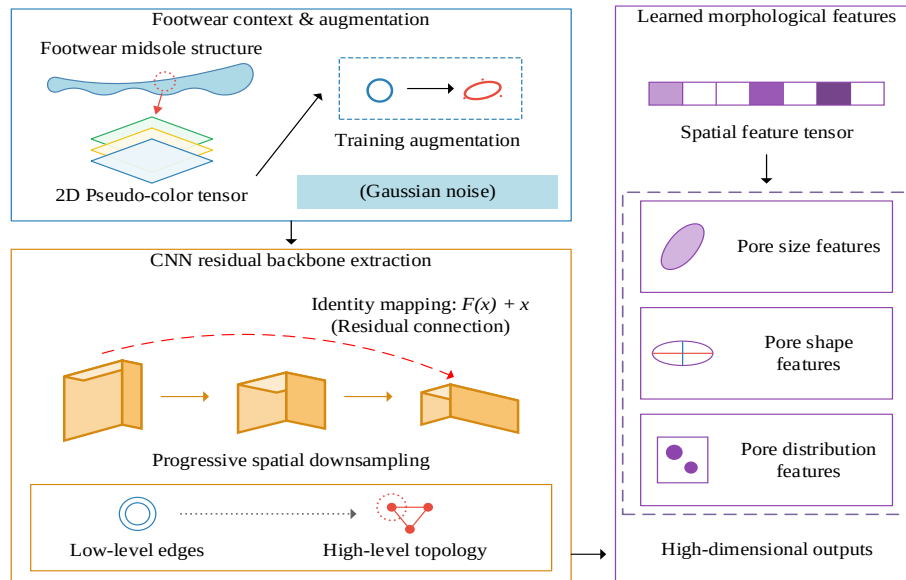


Fig. 3. CNNs downsampling extraction network architecture diagram for micro topological features

Note: The footwear sampling stage illustrates the intended application of the framework.

In Figure 3, Gaussian noise is added to the distance-map inputs during training to represent input perturbations. Subsequently, the two-dimensional pseudo-color tensor enters the base of the

residual network. Relying on the identity mapping and step-by-step spatial down-sampling mechanism, the network deeply abstracts shallow edge features into high-order topological semantics while filtering out matrix noise. The residual identity mapping of micro topological features is shown in Equation (2).

$$Y_{topo} = F(X_{in}, \{W_k\}) + X_{in} \quad (2)$$

In Equation (2), Y_{topo} is the learned spatial feature tensor output by CNNs. X_{in} is the input two-dimensional pseudo-color dimensionality reduction feature matrix. F is the multi-layer nonlinear mapping function in the residual network base. W_k represents the adaptive weight parameter used to extract the edge features of the foaming structure in each convolution layer. The CNN output is retained as a learned spatial feature tensor and passed to the subsequent window-attention blocks for modeling interactions among local morphological features.

2.3 Construction of Durability Prediction Network Integrating Swin Transformer

2.3.1 Modeling of Long-Range Dependence of Pores Based on Self-Attention

Although CNNs have efficiently quantified microscopic local features, the foamed matrix is a highly cross-linked porous skeleton. Cyclic loading can drive microscopic crack propagation and structural failure in heterogeneous materials [17]. Traditional CNNs are limited by their local receptive fields, making it difficult to capture the long-range dependence of stress transfer across space. To break through this bottleneck, a study has been carried out on the interaction center that introduces the self-attention mechanism to reconstruct the model. In terms of algorithm deployment, the model divides the high-dimensional feature tensor output by the CNNs into non-overlapping local windows, and builds a multi-head self-attention calculation graph internally. Successive blocks alternate between regular and shifted window partitions, enabling information exchange across neighboring windows. The spatial weight solution for the long-range dependence of pores is shown in Equation (3).

$$A_{swin} = \text{Softmax} \left(\frac{QK^T}{\sqrt{d_s}} + B_{pos} \right) V \quad (3)$$

In Equation (3), A_{swin} is the attention feature map output by Swin Transformer encoding interactions among image features. Q , K , and V respectively represent the query matrix, key matrix and value matrix generated by linear mapping of local window features. d_s is the dimension of the query and key vectors in each attention head, and its square root scales the dot products before SoftMax normalization. B_{pos} is the learnable relative position bias representing token offsets within each window. Alternating regular and shifted windows enables information exchange across neighboring spatial regions.

2.3.2 Microscopic Pore Feature Enhancement Network Integrating Multi-Dimensional Attention

Although the model has captured long-range dependence, local cell morphology also influences the compressive response of foams [18]. Attention mechanisms can emphasize task-relevant features through adaptive weighting [19]. The study implements three-line parallel analysis of channel and spatial dimensions on the feature tensor, extracts and aggregates complementary weights to reweight the learned morphological features, and the architecture is shown in Figure 4.

In Figure 4, the network uses three-way orthogonal decomposition to independently calculate the self-attention thermal distribution of channels and spatial height and width dimensions. Each complementary weight is then aggregated and non-linearly activated to generate a global mask, which is element-wise multiplied with the original features. The multi-dimensional feature-reweighting mechanism is shown in Equation (4).

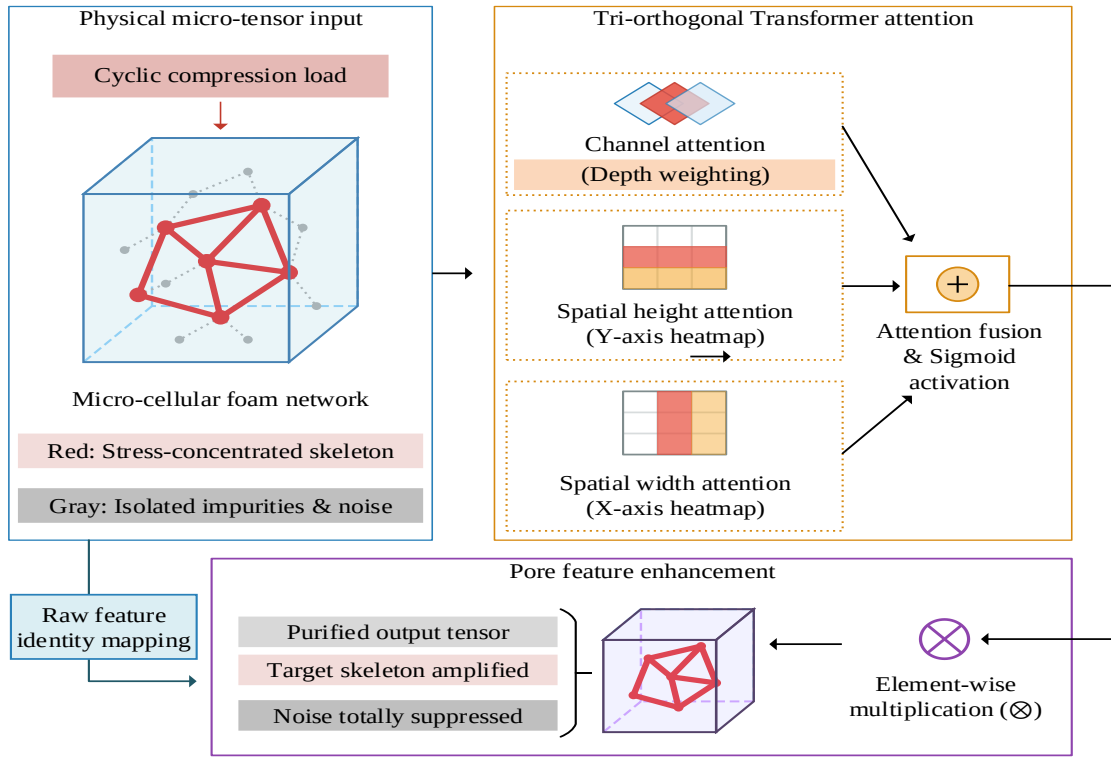


Fig. 4. Pore feature enhancement network diagram fused with multi-dimensional Transformer

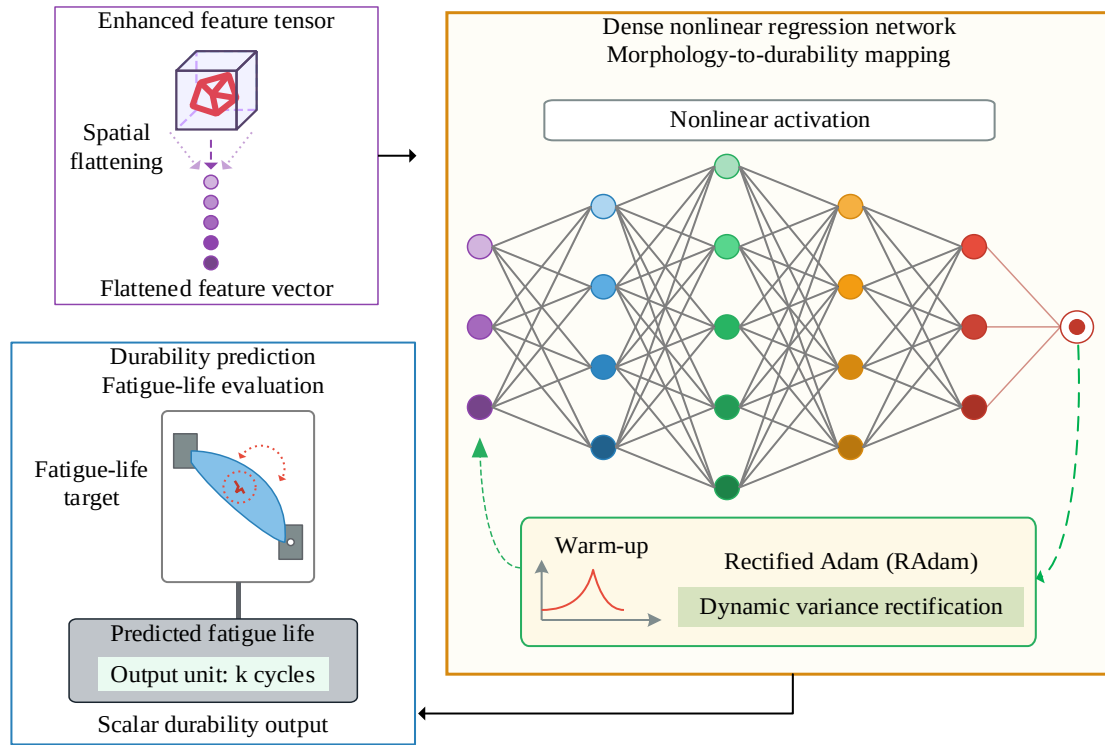


Fig. 5. Fully connected regression network diagram from microscopic pore characteristics to macroscopic durability

$$F_{out} = F_{in} \otimes \sigma(\lambda_c A_c \oplus \lambda_h A_h \oplus \lambda_w A_w) \quad (4)$$

In Equation (4), F_{out} is the enhanced feature tensor output after attention-based reweighting. F_{in} is the original topological feature of the input attention center. $\otimes$ represents the element-wise multiplication operation of the matrix. σ is the *Sigmoid* nonlinear activation function, used to generate a smooth global feature mask. A_c , A_h , and A_w represent the orthogonal attention heat

distribution map of the three dimensions of channel, height, and width respectively. λ is the learnable balance coefficient of the corresponding dimension. $\oplus$ represents the aggregation and addition of multi-dimensional features. The resulting mask adjusts the contributions of learned features before nonlinear regression.

2.3.3 Nonlinear Regression Mapping for Macroscopic Durability Prediction

Neural network models of nonlinear material behavior require suitable architectures and training conditions [20]. Deep neural networks can learn mappings from microstructural images to homogenized mechanical responses [21]. The nonlinear regression network is trained using the Rectified Adam (RAdam) optimization algorithm. The network is shown in Figure 5.

In Figure 5, to bridge the physical gap, the enhanced feature tensor is spatially flattened into a one-dimensional feature vector and directly input into the fully connected regression network. The network uses multiple hidden layers and nonlinear activation functions to fit the mapping from micro topology to the selected durability target. The regression mapping for macro durability indicators is shown in Equation (5).

$$\begin{cases} Z_{topo}^i = \psi_i(\omega_i^T Z_{topo}^{i-1} + b_i), & i = 1, \dots, L \\ P_{life} = \psi(\omega_{out}^T Z_{topo}^L + b_{out}) \end{cases} \quad (5)$$

In Equation (5), Z_{topo}^i is the input vector obtained by flattening the enhanced feature tensor, and P_{life} is the output of hidden layer, where i is the hidden-layer index and L is the total number of hidden layers. The weight matrix, bias vector, and activation function of hidden layer i are denoted by ω_i , b_i , and ψ_i , respectively. The output-layer weight vector and scalar bias are denoted by ω_{out} and b_{out} , and ψ is the output activation function. The superscript T denotes transposition. P_{life} is the scalar prediction for the selected durability target and has the same unit as its reference label. To overcome the gradient oscillation in the early stage of training, the RAdam algorithm is deployed at the bottom layer. Relying on the warm-up and dynamic variance correction mechanism, it effectively ensures the smooth optimization of weights and efficient convergence of the network. The predicted scalar output represents fatigue life under the material and loading conditions assigned to the corresponding numerical case.

3. Results

3.1 Experimental Settings and Configuration

3.1.1 Data Sources and Reference Labels

The experiment relied on the Foamed Cement API RP 10B-4 dataset from Digital Rocks Portal (project DRP-154; DOI: 10.17612/P7GD4Q) and the RealPUFoam volume archived in Zenodo (record 17590658; DOI: 10.5281/zenodo.17590658) for evaluation. The cement dataset contains two formulations prepared with nominal foam qualities of 10%, 20%, 30%, and 40%. The polyurethane foam volume comprises $1300 \times 951 \times 960$ voxels with an isotropic voxel spacing of $2.75 \mu\text{m}$. The selected tomograms provide pore–solid interfaces and spatial variations in cell morphology for evaluating the distance-map representation. Application to footwear foams requires correspondence in pore morphology, connectivity, physical pore scale, bulk density, and material response under the specified loading conditions.

The two sources were therefore used as morphology-oriented porous-media benchmarks rather than as direct substitutes for production footwear-foam datasets. Their relevance lies in providing resolved pore–solid interfaces, connectivity variations, and porosity ranges required to evaluate the distance-map representation and morphology-to-durability mapping. Differences may nevertheless remain relative to industrial EVA or PU foams in polymer formulation, cell-size distribution, anisotropy, bulk density, and processing history. Transfer to footwear materials therefore requires material-specific calibration and independent validation.

Reference fatigue-life labels were obtained from numerical analyses of the reconstructed geometries. Each label was linked to its source volume, derived geometry, material-parameter record, and loading history through persistent identifiers.

3.1.2 Specimen Partitioning and Image Preprocessing

Parent specimens were assigned to the training, validation, and test subsets in a 70:15:15 ratio using a fixed split seed of 2026. All slices, crops, augmented inputs, and derived geometries originating from the same parent specimen were retained in the same subset. Model selection used the validation subset, and the test subset was reserved for final evaluation.

For each labeled geometry, three orthogonal slice triplets were extracted at relative positions of 25%, 50%, and 75% along each axis. Each slice was centrally cropped to a square and resized to 224×224 pixels using bilinear interpolation. Intensities were clipped at the slice-specific first and ninety-ninth percentiles and rescaled to $[0, 1]$. Contrast-limited adaptive histogram equalization used 32×32 -pixel regions, a clip limit of 0.01, and 256 bins. Li's minimum cross-entropy thresholding used the mean intensity for initialization and a tolerance of 10^{-5} ; pixels below the threshold represented pores.

Euclidean distance maps, with pore pixels set to zero, were divided by the image diagonal (315.37 pixels) and stacked into three-channel inputs. Training augmentation used zero-mean Gaussian noise with a standard deviation of 0.01 and clipping to $[0, 1]$. Validation and test inputs were not augmented. Predictions from the three slice triplets were averaged to obtain one prediction for each labeled geometry.

3.1.3 Model Configuration and Training

To objectively measure the regression accuracy of the proposed method, the experiment selected four visual architectures as comparison benchmarks: ResNet-50, ViT-Base, ConvNeXt-V2, and FaViT-B2 [9]. FaViT-B2 was included as a recent Transformer-based reference for long-range dependency modeling. Each architecture was adapted to scalar fatigue-life regression and evaluated under the same specimen partition, input representation, training budget, and random-seed protocol. All experiments were executed independently in the PyTorch environment on an Ubuntu workstation equipped with an RTX 4090 graphics card.

The CNN comprised two stages with 48 and 96 channels, respectively. Each stage used a 3×3 stride-2 convolution, batch normalization, and ReLU, followed by two basic residual blocks with two 3×3 stride-1 convolutions per block. All convolutions used padding of one pixel. The output tensor was $96 \times 56 \times 56$. The Swin component used four stages with channel widths of 96, 192, 384, and 768; depths of 2, 2, 6, and 2; and attention-head counts of 3, 6, 12, and 24, respectively. The CNN output entered the first Swin stage directly. Patch merging halved spatial resolution and doubled channel width between stages. The 7×7 windows alternated between shifts of zero and three tokens, with no shifting in the final 7×7 stage. The feed-forward expansion ratio was 4, and the maximum stochastic-depth probability was 0.2.

The enhanced $768 \times 7 \times 7$ tensor was flattened into a 37,632-dimensional vector. Four hidden layers with 256, 128, 64, and 32 units each used ReLU and dropout of 0.1, followed by a linear scalar output.

Training used RAdam with $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$, and a coupled L2 weight-decay coefficient of 10^{-4} . The learning rate increased linearly from 10^{-6} to 10^{-4} during the first five epochs and subsequently followed cosine decay to 10^{-6} at epoch 100. The batch size was 32. Fatigue-life labels were converted from cycles to k cycles by division by 1,000, without additional target standardization. The training objective was mean squared error in the converted target unit. Training continued for 100 epochs, and the checkpoint with the lowest validation RMSE was retained for evaluation.

3.1.4 Numerical Simulation and Loading Conditions

Fatigue life was defined as the first cycle at which the accumulated fatigue damage reached 1.0. Damage accumulation followed the Palmgren–Miner rule using a calibrated stress–life relationship for the assigned solid-phase material. The six numerical screening cases had porosities of 52.4%, 58.7%, 64.1%, 71.5%, 76.8%, and 82.3%, respectively. Prediction errors were evaluated at impact energies of 2, 4, 6, and 8 J. The reported compression-response trajectory covered 100,000 loading cycles.

3.1.5 Statistical Analysis and Inference Timing

Each model was evaluated over ten training runs using random seeds from 2026 to 2035 and the same parent-specimen partition. Absolute and squared errors were first averaged within each parent specimen; MAE was the mean parent-level absolute error, and RMSE was the square root of the mean parent-level squared error. Performance across training runs was summarized as mean $\pm$ sample standard deviation.

Pointwise 95% confidence intervals for the run-averaged RMSE and MAE used 10,000 bootstrap resamples of test parent specimens, retaining all geometries of each selected parent and using identical indices across models. Metrics were recalculated per run and then averaged across runs; the 2.5th and 97.5th percentiles defined intervals conditional on the fitted models.

MAE comparisons used two-sided paired permutation tests on parent-level absolute errors averaged across runs, with 9,999 random swaps or exact enumeration when fewer distinct swaps existed. Holm correction covered four planned comparisons between IPM-FMD and the four baseline models at a family-wise significance level of 0.05.

Network-only latency was measured in evaluation mode without gradients, using GPU-resident FP32 inputs and a batch size of one. Each of the ten checkpoints underwent 100 warm-up passes and 1,000 timed forward passes using CUDA events and synchronization. Latency was averaged per checkpoint and summarized across checkpoints. Preprocessing, host-to-device transfer, and multi-slice aggregation were excluded and timed separately for complete-pipeline evaluation.

3.2 Performance Comparison Verification of IPM-FMD

3.2.1 High-Precision Regression Verification of The Fatigue Bending Resistance Limit of Foaming Media

To quantitatively evaluate the error interval and feature mapping accuracy of the architectures visualized in Figure 6 for predicting the fatigue bending life of foaming media, the experiment comprehensively examined the regression accuracy and computing power overhead of the model, as shown in Figure 6.

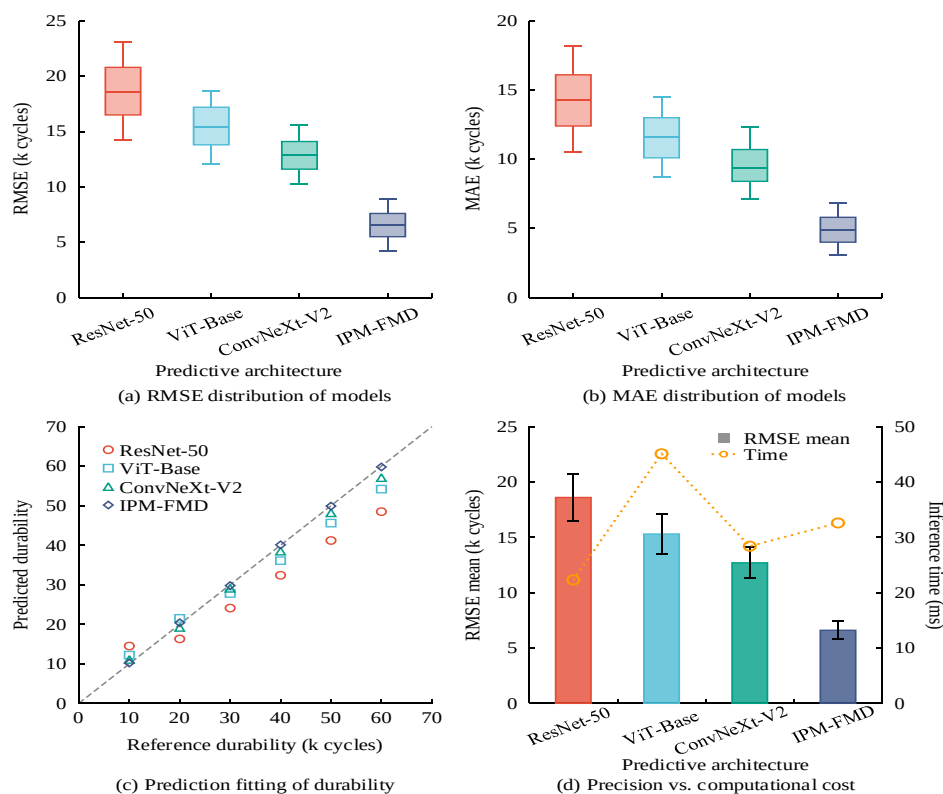


Fig. 6. Comprehensive performance comparison of various prediction architectures in the foaming medium durability regression task

In Figures 6(a) and 6(b), IPM-FMD shows lower and more compact RMSE and MAE distributions than the baseline architectures visualized in the figure. Its RMSE distribution is concentrated around 6.6 k cycles, reflecting stable prediction performance. Figure 6(c) further shows that the predicted values of IPM-FMD remain close to the ideal diagonal across the evaluated fatigue-life range. In contrast, ResNet-50 underestimates the reference lifetime at 60 k cycles, with an absolute prediction error of up to 11.5 k cycles. From Figure 6(d), IPM-FMD reduces the mean RMSE to 6.60 k cycles, while its network-only inference time of 32.6 ms remains lower than the 45.1 ms of ViT-Base, providing a favorable balance between regression accuracy and computational cost.

To further quantify the performance variability of different prediction architectures under repeated training, ResNet-50, ViT-Base, ConvNeXt-V2, FaViT-B2, and IPM-FMD were evaluated over ten independent training runs under the same Digital Rocks Portal and Zenodo data protocol, specimen partition, and training settings. Table 1 summarizes RMSE and MAE as mean $\pm$ standard deviation with 95% confidence intervals, together with network inference time and Holm-adjusted paired-comparison results.

Table 1. Statistical comparison of IPM-FMD and baseline models over ten independent runs

Model	RMSE, mean $\pm$ SD (95% CI), k cycles ↓	MAE, mean $\pm$ SD (95% CI), k cycles ↓	Inference time, mean $\pm$ SD (ms) ↓	Holm-adjusted <i>p</i> - value
ResNet-50	18.60 $\pm$ 2.72 (16.50–20.70)	14.31 $\pm$ 2.48 (12.53–16.09)	22.3 $\pm$ 0.6	0.004
ViT-Base	15.30 $\pm$ 2.12 (13.50–17.10)	11.59 $\pm$ 1.88 (10.25–12.93)	45.1 $\pm$ 1.4	0.007
ConvNeXt-V2	12.70 $\pm$ 1.72 (11.30–14.10)	9.55 $\pm$ 1.63 (8.38–10.72)	28.4 $\pm$ 0.8	0.016
FaViT-B2	10.85 $\pm$ 1.70 (9.63–12.07)	8.16 $\pm$ 1.28 (7.25–9.07)	36.8 $\pm$ 1.4	0.031
IPM-FMD	6.60 $\pm$ 1.53 (5.80–7.40)	4.92 $\pm$ 1.18 (4.08–5.76)	32.6 $\pm$ 1.0	/

Note: Values are reported as mean $\pm$ standard deviation over ten independent training runs; values in parentheses denote 95% confidence intervals. Adjusted *p*-values refer to paired MAE comparisons with IPM-FMD after Holm correction. Lower RMSE, MAE, and inference time indicate better performance.

As shown in Table 1, IPM-FMD achieved the lowest mean RMSE and MAE, reaching 6.60 $\pm$ 1.53 k cycles and 4.92 $\pm$ 1.18 k cycles, respectively. Compared with FaViT-B2, these errors were reduced by approximately 39.2% and 39.7%. The Holm-adjusted *p*-values for the paired MAE comparisons between IPM-FMD and the four baseline models were all below the predefined 0.05 significance level. In terms of computational efficiency, ResNet-50 and ConvNeXt-V2 required shorter inference times of 22.3 and 28.4 ms, respectively, whereas their prediction errors were substantially higher. IPM-FMD required 32.6 ms per network forward pass, which remained lower than FaViT-B2 at 36.8 ms and ViT-Base at 45.1 ms. Therefore, the proposed model mainly provides a favorable balance between prediction accuracy and computational cost rather than the lowest latency alone. Inference time represents network-only batch-1 latency on the RTX 4090 GPU; the detailed timing and statistical procedures are described in Section 3.1.5.

3.2.2 Engineering Generalization and Fault Tolerance Under Extreme Pore Distortion Scenarios

To examine the fault-tolerance performance and engineering generalization robustness of the IPM-FMD model when dealing with severe distortion of the pores of the foaming medium, the experiment further dynamically tracked the training convergence process of the architectures visualized in Figure 7, and objectively analyzed its error fluctuation trend within the differentiated porosity interval, as shown in Figure 7.

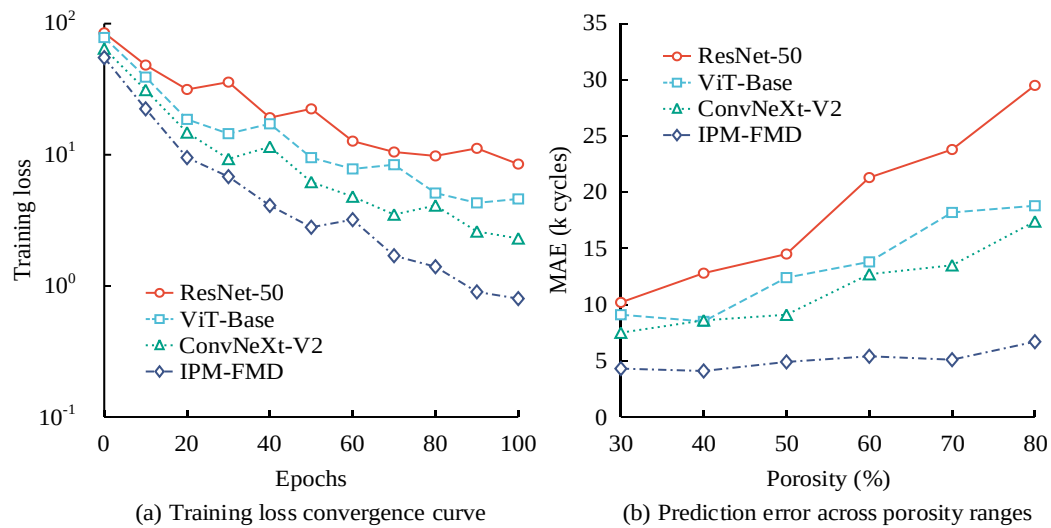


Fig. 7. Convergence stability and pore distortion tolerance performance evaluation of the prediction model under extreme working conditions

In Figure 7(a), affected by stochastic gradient perturbation, the benchmark model experienced significant gradient jumps during the convergence process. Although the IPM-FMD model produced slight oscillations at 60 rounds, the final Loss value dropped steadily to 0.8, demonstrating extremely strong training stability. As shown in Figure 7(b), as the porosity evolves to the extreme distortion interval of 80%, the prediction error MAE of ResNet-50 sharply soars to 29.5 k cycles, showing a serious generalization bottleneck. In contrast, the IPM-FMD model maintains lower prediction errors across the evaluated porosity range, with an MAE of 5.1 k cycles at 70% porosity. These results indicate improved robustness to pore-distortion variation under the benchmark conditions, while transfer to real footwear foams requires independent validation.

3.3 Ablation Experimental Analysis of IPM-FMD Model

To further verify the independent contribution of feature enhancement and other modules in the IPM-FMD model to durability prediction, the experiment further established the ablation scheme. Three degenerate variants are studied: No Multi-Attention (w/o MDA), No Swin-Transformer (w/o ST) and Standard Adam (Adam). The test environment and data follow the configuration of Digital Rocks Portal and Zenodo data set. The specific ablation experiment variable control and configuration details are shown in Table 2.

In Figure 8(a), the loss value of the complete IPM-FMD model quickly dropped to 1.15 after the 40th round, and finally converged to 0.81, while the w/o ST and Adam variants oscillated significantly due to the influence of topological noise. As shown in Figure 8(b), after stripping off the multidimensional attention or window transformer, the RMSE of its fatigue life prediction climbs to 8.8 k cycles and 12.5 k cycles respectively, indicating that long-range modeling contributes to controlling extreme deviations. As shown in Figure 8(c), the median absolute error of the complete model is only 5.4 k cycles and the interquartile range is the narrowest, supporting the combined contribution of the core mechanisms to system robustness when coping with extreme deformation.

Because each input channel encodes normalized distance-to-pore information, the learned spatial features remain directly related to pore proximity and local skeleton geometry. The increased prediction errors after removing the multi-dimensional attention or Swin Transformer indicate that spatial reweighting and non-local feature interaction both contribute to fatigue-life regression. However, these ablations do not localize the contribution of individual pore regions, and no dedicated post-hoc saliency analysis was performed in the present study. Future validation will therefore combine attention-rollout or saliency mapping with pore-size, shape, connectivity, and local stress descriptors to examine whether high-contribution regions correspond to physically critical microstructures.

Table 2. Ablation experiment variable control and configuration of IPM-FMD model

Model/Variant	CNNs Micro-topology	Swin Transformer	Multi-dimensional Attention	RAdam Optimizer	Configuration Description
IPM-FMD (Full)	✓	✓	✓	✓	Research model retaining all core mechanisms
IPM-FMD (w/o MDA)	✓	✓	×	✓	Removal of multi-dimensional feature reinforcement
IPM-FMD (w/o ST)	✓	×	✓	✓	Swin Transformer replaced by deep residual convolution
IPM-FMD (Adam)	✓	✓	✓	×	Regressor optimization degraded to standard Adam

In Table 2, the experiment strips off the multi-dimensional attention module at fixed points to test the feature purification performance, downgrades the window transformer to depth convolution to evaluate the long-range dependency capture gain, and degrades the optimizer to a standard algorithm to examine the convergence smoothness. The comprehensive regression performance and dynamic error results of each model based on this setting are shown in Figure 8.

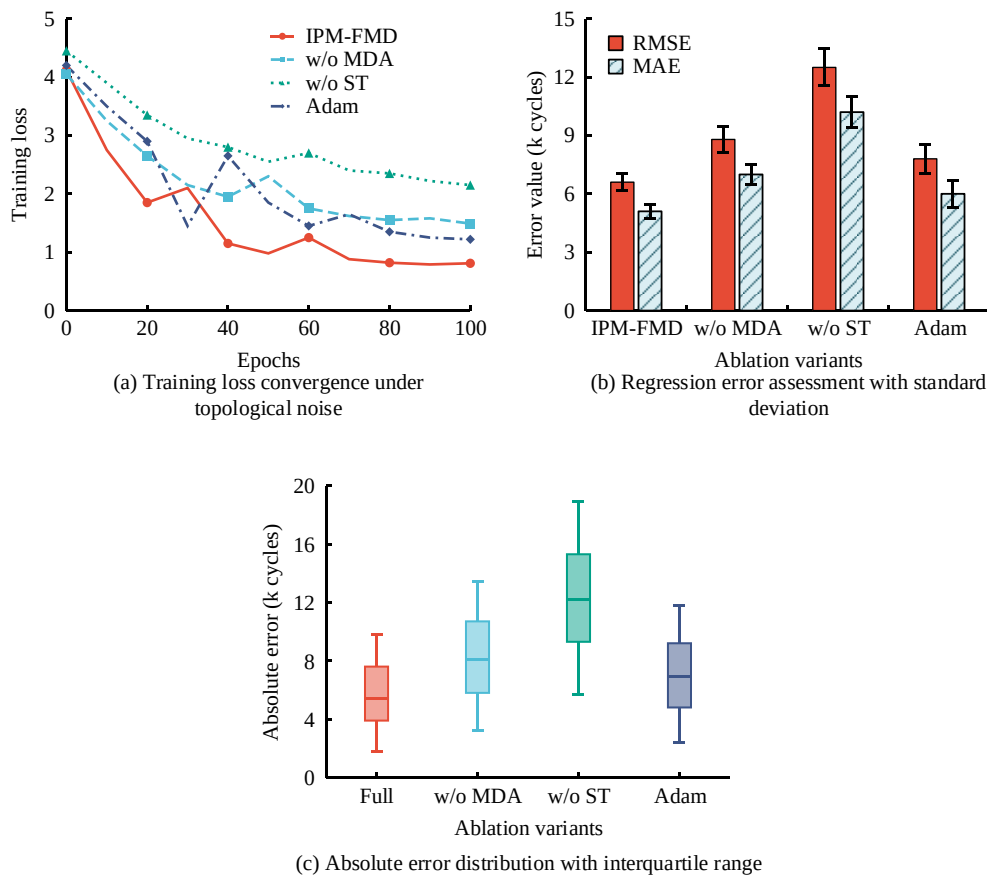


Fig. 8. Comprehensive performance comparison of ablation experiments of core components of IPM-FMD model

3.4 Analysis of Simulation Application Effects of IPM-FMD

3.4.1 Micro-Topology Directional Optimization for Fatigue-Resistant Design

Focusing on the evaluation bottleneck of shoe material foam structures in real engineering implementation, the experiment followed the Digital Rocks Portal and Zenodo open-source data sets to expand the prediction boundary to dynamic physical simulation scenarios. To bridge the mechanical gap between theoretical models and complex service conditions, the core data mapping logic was extended from single static topography analysis to multi-modal stress environment and high-throughput concurrent optimization of feature fields. To verify the topology optimization effectiveness of the predictive feedback mechanism in the reverse design of foaming formulas, experiments were conducted to quantitatively compare the anti-fatigue mechanical performance of the initial random matrix and the pore characteristics after iterative reconstruction, as shown in Figure 9.

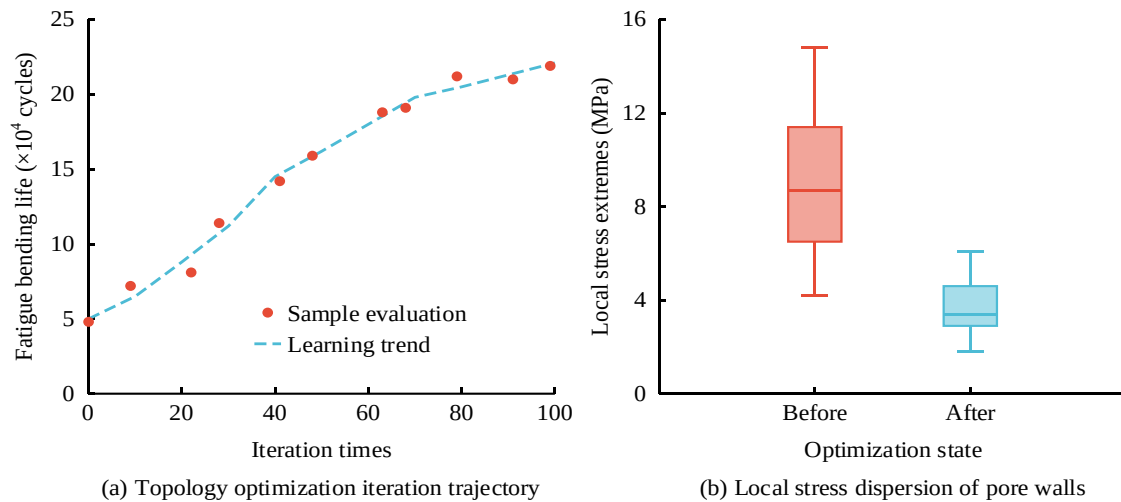


Fig. 9. Comparison of microtopology optimization trajectory and local stress dispersion based on IPM-FMD

In Figure 9(a), the topology optimization trajectory of IPM-FMD prediction feedback shows a significant nonlinear climbing trend. The initial matrix fatigue life was only 4.8×10^4 cycles. After about 80 iterations, this indicator exceeded 21.2×10^4 times and dynamically converged, verifying the efficient directional driving force of the model. As shown in Figure 9(b), before optimization, the local stress extremes of the pore wall were highly divergent, with a median of 8.7MPa and a maximum of 14.8MPa, indicating severe stress concentration. After optimization, not only did the median stress drop to 3.4MPa, but the interquartile range also narrowed significantly, with the highest extreme being controlled at 6.1MPa, which intuitively confirmed the model's powerful ability to repair microscopic defects.

3.4.2 Validation of High-Throughput Virtual Screening Performance of Massive Foaming Formulas

To verify the computational efficiency and engineering substitution potential of the IPM-FMD model in the concurrent evaluation scenario of massive recipes, experiments were carried out on six groups of representative pore gradient simulation schemes to quantitatively compare the time consumption and accuracy with traditional finite element analysis (FEA), as shown in Table 3.

In Table 3, the study quantitatively compared the screening effectiveness of the IPM-FMD model and traditional FEA on porous samples. When the porosity rises to 82.3%, the traditional FEA solution time increases to 24.6 hours, while the IPM-FMD virtual-screening time remains within 2.4 seconds, achieving an order-of-magnitude improvement in efficiency. In terms of accuracy, even in the face of extremely distorted pores, its predicted life is still highly consistent with the benchmark, and the maximum relative error is strictly controlled within 5.54%. This fully confirms that this model has extremely high engineering substitution potential in virtual concurrent screening of massive materials.

Table 3. Comparison of performance and accuracy between IPM-FMD model and traditional FEA in high-throughput simulation screening

Simulation Test Group	Equivalent Porosity (%)	Traditional FEA Time (h)	IPM-FMD Screening Time (s)	FEA Baseline Predicted Life ($\times 10^4$ cycles)	IPM-FMD Model Predicted Life ($\times 10^4$ cycles)	Absolute Relative Error (%)
Sample Group A	52.4	14.5	2.1	18.42	18.15	1.46
Sample Group B	58.7	15.2	2.2	15.75	16.02	1.71
Sample Group C	64.1	16.8	2.1	12.30	12.05	2.03
Sample Group D	71.5	18.5	2.3	8.95	9.24	3.24
Sample Group E	76.8	21.4	2.4	6.42	6.18	3.74
Sample Group F	82.3	24.6	2.4	4.15	4.38	5.54

3.4.3 Fatigue Failure Prediction and Tracking Under Complex Gait Impact Conditions

To verify the predictive robustness of the IPM-FMD model in the face of complex dynamic loads such as gait transitions, experiments quantified its dual error evolution under multi-level load transitions and tracked its long-term fatigue attenuation fit trajectory with the FEA benchmark, as shown in Figure 10.

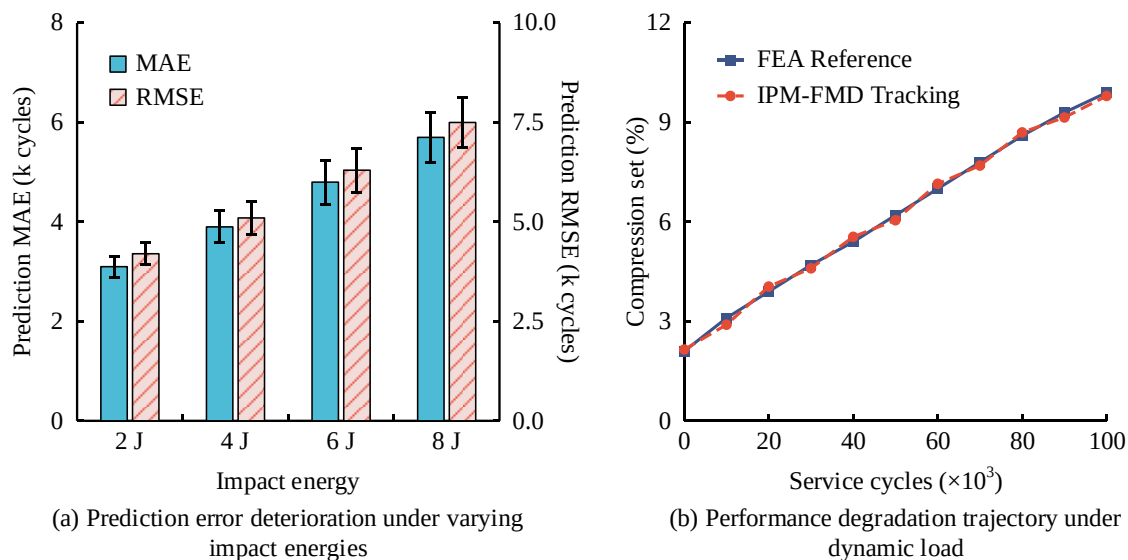


Fig. 10. Error evolution and fatigue attenuation tracking of the IPM-FMD model under variable load impact conditions

In Figure 10(a), under increasing dynamic load, the prediction error of the IPM-FMD model increases in a controllable manner. When the impact energy jumped from 2J to 8J, the maximum RMSE was still strictly controlled at 7.50 k cycles, and the maximum MAE was only 5.70 k cycles, demonstrating excellent generalization robustness in the face of complex load mutations. As shown in Figure 10(b), in one hundred thousand continuous simulation tracking, the compression deformation trajectory of the research model was highly consistent with the FEA benchmark. Even at the end of service, the predicted value (9.80%) remained close to the FEA reference value (9.90%), indicating close agreement with the numerical benchmark over the simulated loading trajectory. From an application perspective, the 32.6 ms result represents network-only batch-1 inference rather than the complete micro-CT acquisition and processing workflow. Although this computational cost supports rapid virtual material screening, it does not by itself demonstrate real-

time in-line quality control. The present evaluation also does not include real commercial footwear-foam micro-CT volumes from independent manufacturing batches. Variations in EVA or PU formulation, density, foaming and curing conditions, production batch, aging state, and loading environment may shift the morphology–durability relationship observed in the benchmark data. Accordingly, practical deployment requires material-specific calibration and external validation using independent footwear-foam batches before the model is incorporated into production screening or quality-control workflows.

4. Conclusions

To solve the problem that the mapping rules between the microscopic pore characteristics and macroscopic durability of shoe material foaming media are difficult to reveal and that traditional testing is expensive, an IPM-FMD model was constructed. This model relies on CNNs to quantify micro topology, and integrates Swin Transformer and multi-dimensional attention to accurately capture the long-range dependence of stress and purify the load-bearing skeleton, achieving high-precision nonlinear regression of macroscopic fatigue life. The results show that the median predicted RMSE is only 6.6 k cycles, and the generalization is stable under 80% extreme pore distortion. Across ten independent runs, the model achieved a mean RMSE of 6.60 ± 1.53 k cycles and a mean MAE of 4.92 ± 1.18 k cycles, while the network-only batch-1 inference latency was 32.6 ± 1.0 ms on the RTX 4090. In engineering simulation applications, model-driven micro-optimization allows the fatigue life to exceed 21.2×10^4 cycles, while the virtual-screening time remained within 2.4 s per sample, compared with up to 24.6 h for traditional finite element analysis.

Although the prediction performance is excellent, the current two-dimensional cross-section dimensionality reduction strategy is prone to losing cross-force information when dealing with complex three-dimensional connected networks. The present validation is also based on porous-media benchmark volumes and numerically generated fatigue-life references rather than independent production batches of commercial footwear foams. Differences in EVA or PU formulation, density, foaming and curing conditions, manufacturing batches, aging state, and loading environments may alter the morphology–durability relationship. Moreover, the reported 32.6 ms latency represents network-only inference rather than the complete micro-CT acquisition and preprocessing workflow. These factors limit direct extrapolation of the current results to industrial in-line quality control. In view of this limitation, follow-up research will introduce cutting-edge algorithms such as 3D point cloud reconstruction to complete the direct analytical closed loop of full 3D topology, and couple multi-modal micromechanics in-situ tracking technology to capture the microscopic stress evolution of the foamed skeleton in real time, thereby driving the intelligent prediction model to expand to more complex and realistic working conditions. Future validation will additionally incorporate real footwear-foam micro-CT data from independent manufacturing batches and examine formulation, aging, and loading variations. Attention-rollout or saliency analysis will also be introduced to relate model sensitivity to physically meaningful pore characteristics before deployment in material-screening or quality-control workflows.

Fundings

The research is supported by 2024 Annual Regular Project of Fujian Provincial Education Science Planning, Strategic Analysis of International Exchange and Cooperation in Higher Vocational Colleges under the Background of the Belt and Road Initiative, Project No.: FJJKGZ24-018. 2024 Annual Putian Science and Technology Plan Project, A Plasma Technology-based Process for Changing the Surface Treatment of Shoe Materials in Traditional Shoe Manufacturing, Funding No.: GM202403. 2024 Putian Science and Technology Plan Project, Construction of Putian Shoe Industry Technology Innovation Research Institute, Funding No.: 2024HJGZ001.

References

- [1] Parente JM, Reis PNB, Berto F. Fatigue response of auxetic structures: A review. *Fatigue & Fracture of Engineering Materials & Structures*. 2024;47(8):2679-2699. <https://doi.org/10.1111/ffe.14351>

- [2] Hoffmann M, Pai SM, Albuquerque RQ, Kastl S, Ruckdäschel H. Active learning-driven inverse design of polyurethane foams for EV battery applications. *Journal of Polymer Science*. 2025;63(21):4621-4630. <https://doi.org/10.1002/pol.20250250>
- [3] Herrmann L, Kollmannsberger S. Deep learning in computational mechanics: a review. *Computational Mechanics*. 2024;74(2):281-331. <https://doi.org/10.1007/s00466-023-02434-4>
- [4] Nguyen DM, Zhang J, Rahim M, Hoang D, Promis G, Ganaoui ME, et al. A review of microstructural characterization and liquid transport in porous materials through image processing techniques. *Transport in Porous Media*. 2024;151(7):1643-1664. <https://doi.org/10.1007/s11242-024-02088-7>
- [5] Zhuang W, Wang E, Zhang H. Prediction of the compressive mechanical properties and reverse structural design of two-dimensional mesoscopic aluminum foam based on deep learning methods. *Journal of Materials Science*. 2024;59(25):11416-11439. <https://doi.org/10.1007/s10853-024-09866-0>
- [6] Kadulkar S, Sherman ZM, Ganesan V, Truskett TM. Machine learning-assisted design of material properties. *Annual Review of Chemical and Biomolecular Engineering*. 2022;13(1):235-254. <https://doi.org/10.1146/annurev-chembioeng-092220-024340>
- [7] Khan A, Rauf Z, Sohail A, Khan AR, Asif H, Asif A, et al. A survey of the vision transformers and their CNN-transformer based variants. *Artificial Intelligence Review*. 2023;56(Suppl 3):2917-2970. <https://doi.org/10.1007/s10462-023-10595-0>
- [8] Alrfou K, Zhao T, Kordijazi A. Deep learning methods for microstructural image analysis: the state-of-the-art and future perspectives. *Integrating Materials and Manufacturing Innovation*. 2024;13(3):703-731. <https://doi.org/10.1007/s40192-024-00369-z>
- [9] Qin H, Zhou D, Xu T, Bian Z, Li J. Factorization vision transformer: Modeling long-range dependency with local window cost. *IEEE Transactions on Neural Networks and Learning Systems*. 2025;36(2):3151-3164. <https://doi.org/10.1109/TNNLS.2023.3342172>
- [10] Sahoo MK, Mandal A. Three-dimensional design and simulation of carbon foam. *Journal of Cellular Plastics*. 2025;61(2):103-134. <https://doi.org/10.1177/0021955X251315467>
- [11] Andrä H, Dobrovolskij D, Engelhardt M, Godehardt M, Makas M, Mercier C, et al. Image-based microstructural simulation of thermal conductivity for highly porous wood fiber insulation boards: 3D imaging, microstructure modeling, and numerical simulations for insight into structure-property relation. *Wood Science and Technology*. 2023;57(1):5-31. <https://doi.org/10.1007/s00226-022-01434-6>
- [12] Prifling B, Weber M, Ray N, Prechtel A, Phalempin M, Schlüter S, et al. Quantifying the impact of 3D pore space morphology on soil gas diffusion in loam and sand. *Transport in Porous Media*. 2023;149(2):501-527. <https://doi.org/10.1007/s11242-023-01971-z>
- [13] Seibert P, Raßloff A, Kalina K, Kästner M. Fast descriptor-based 2D and 3D microstructure reconstruction using the Portilla-Simoncelli algorithm. *Engineering with Computers*. 2025;41(1):589-607. <https://doi.org/10.1007/s00366-024-02026-7>
- [14] Raghavendra S, Russo F, De Biasi R, Rustighi E, Zappini G, Berto F, et al. Characterization of compressive fatigue behavior and acoustic emission analysis of Ti6Al4V cellular lattice materials fabricated by laser powder bed fusion. *Fatigue & Fracture of Engineering Materials & Structures*. 2024;47(10):3910-3929. <https://doi.org/10.1111/ffe.14387>
- [15] Gorgularslan RM, Ozay AI. Fatigue life prediction of lattice structures using a strain-life method based on an equivalent solid model. *Proceedings of the Institution of Mechanical Engineers, Part L: Journal of Materials: Design and Applications*. 2023;237(7):1548-1561. <https://doi.org/10.1177/14644207221148854>
- [16] Müller-Pabel M, Meuchelböck J, Koch I, Ruckdäschel H, Gude M. On the determination of viscoelastic properties of EPP foam in dependence on pre-strain and loading direction. *Journal of Cellular Plastics*. 2024;60(2):81-96. <https://doi.org/10.1177/0021955X231224770>
- [17] Wingender D, Balzani D. Simulation of crack propagation through voxel-based, heterogeneous structures based on eigenerosion and finite cells. *Computational Mechanics*. 2022;70(2):385-406. <https://doi.org/10.1007/s00466-022-02172-z>
- [18] Meyer zu Reckendorf I, Gabrion X, Bergeret A, Lacoste C, Perrin D, Redko V, et al. Compressive strength properties related to microstructure of chemically blown tannin foams. *Polymer Engineering & Science*. 2025;65(8):4286-4300. <https://doi.org/10.1002/pen.27295>
- [19] Soydaner D. Attention mechanism in neural networks: where it comes and where it goes. *Neural Computing and Applications*. 2022;34(16):13371-13385. <https://doi.org/10.1007/s00521-022-07366-3>
- [20] Rosenkranz M, Kalina KA, Brummund J, Kästner M. A comparative study on different neural network architectures to model inelasticity. *International Journal for Numerical Methods in Engineering*. 2023;124(21):4802-4840. <https://doi.org/10.1002/nme.7319>
- [21] Aldakheel F, Elsayed ES, Zohdi TI, Wriggers P. Efficient multiscale modeling of heterogeneous materials using deep neural networks. *Computational Mechanics*. 2023;72(1):155-171. <https://doi.org/10.1007/s00466-023-02324-9>