Textile & Leather Review

ISSN 2623-6281| www.tlr-journal.com | 10.31881/TLR

Interfacial Bonding Properties of Natural Fiber Concrete Using Vision Transformer

Chunmei Yao, Xiangru Dong

How to cite: Yao C, Dong X. Interfacial Bonding Properties of Natural Fiber Concrete Using Vision Transformer. Textile & Leather Review. 2026; 9:1333-1357. https://doi.org/10.31881/TLR.2026.1333

How to link: https://doi.org/10.31881/TLR.2026.1333

Published: 29 April 2026

Interfacial Bonding Properties of Natural Fiber Concrete Using Vision Transformer

Chunmei Yao, Xiangru Dong*

Department of Civil Engineering, Architectural Engineering Institute, Anhui Technical College of Water Resources and Hydroelectric Power, Hefei 231603, Anhui, China

*dongxiangru002@126.com

Article

https://doi.org/10.31881/TLR.2026.1333

Received 28 August 2025; Accepted 13 November 2025; Published 29 April 2026

ABSTRACT

This study addresses the challenge of predicting the interfacial bond strength in natural fiber concrete by analyzing microscopic images. Traditional methods struggle to establish a reliable link between fine-scale interface characteristics and overall performance, especially when multiple types of fibers are involved. We develop a novel approach based on a Vision Transformer model to analyze scanning electron and optical micrographs of interfaces involving jute, sisal, coconut coir, flax, and hemp fibers. The image analysis framework is designed to be robust to variations among different fibers, enhancing its ability to generalize. Experimental results show that the model achieves accurate bond strength predictions on the test set, with a mean absolute error of 0.095, a root mean square error of 0.140, and a coefficient of determination of 0.975. The model maintains low prediction errors even for fiber types not included during training, demonstrating strong generalization. Analysis of the model's focus confirms that it identifies physically meaningful features at the fiber-matrix interface, with coconut coir fiber showing the least interfacial activity, correlating with its measured strength. This work provides an intelligent and interpretable tool for studying and optimizing natural fiber concrete.

KEYWORDS

natural fiber concrete, vision transformer, bond strength prediction, domain adaptation, microscopic images

INTRODUCTION

Natural fiber concrete has gained attention in the field of sustainable construction due to its low-carbon, environmentally friendly nature, widespread availability, and enhanced toughness. Interfacial bonding, as an important parameter measuring the synergistic effect between fiber and matrix [1,2], directly affects the

crack resistance, impact resistance, and long-term service performance of concrete. Traditional testing methods rely on macroscopic mechanical tests to determine bond strength [3-5]. While these methods provide information on overall performance, they fail to reveal the quantitative relationship between microscopic features, such as fiber surface morphology and interfacial transition zone structure, and bond performance. This lack of macroscopic-microscopic correlation limits the effectiveness of concrete mix optimization and fiber modification strategies. Furthermore, different natural fibers exhibit significant differences in surface roughness, pore distribution, and hydrophilicity, making it difficult to accurately predict bond strength across fiber types using a single model [6-8].

This paper focuses on the interfacial bonding properties between various textile-derived natural fibers and cement matrices, involving various fiber types. Marawan Saad's team [9] studied the effect of using natural fiber waste to improve the brittle behavior of HSC (High Strength Concrete). S. Ramu et al. [10] conducted rigorous mechanical tests to compare the effects of alkaline and non-alkaline treated composites, revealing the continuous failure area, area and matrix structure of alkaline and untreated composites. Bahrum Prang Rocky et al. [11] studied four types of bamboo, natural bamboo fibers from each type of bamboo, commercial bamboo viscose fibers, and other conventional fibers to determine and compare their elemental composition, chemical bonding, and behavior. Xindang He et al. [12] conducted a comprehensive review of the non-contact full-field optical measurement technology of digital image correlation (DIC), emphasizing its future applications.

In a related study, Suresh Poyil Subramanyam et al. [13] prepared woven areca sheath fibers (ASF) with epoxy composites of different fiber contents and tested their tribological response and mechanical properties on a three-body wear tester. The effect of fiber content on various properties was also investigated. Rupesh Kumar Tipu's team [14] explored the application of machine learning (ML) models in predicting the compressive, flexural, and splitting tensile strengths of concrete with partial replacement of coarse aggregate with coconut shell. Han Xu's [15] team applied machine learning methods to the identification of microstructures. Through transfer learning and feature visualization, they established a highly accurate and interpretable model based on a small experimental dataset, allowing the model results to be interpreted from a physical and chemical perspective. This work provides a new approach to the identification of microstructures and helps further promote the intelligent research and development of polymers [15]. Lais Kohan's team [16] evaluated how

the fabric weave structure and yarn geometry affect the interaction between two different jute fabrics when used as reinforcement materials in a mortar matrix. Adane Dagnaw Gudayu et al. [17] studied the modification of NF (Natural Fiber) and cement matrix as a method to improve the compatibility and degradation and the performance and life of cement-based materials reinforced with natural fibers. It is effective for a single fiber type or a single data domain, but when faced with cross-domain prediction tasks for multiple types of natural fibers, insufficient feature expression and inter-domain distribution differences remain the core bottleneck [18,19], hindering the generalization ability of the model. This work introduces an analytical framework that fundamentally diverges from traditional methods by resolving the multiscale characterization challenge through global context modeling and cross-domain feature alignment. Traditional image analysis and convolutional neural networks struggle to capture long-range dependencies within the fiber-matrix interface, while conventional machine learning models face limitations in generalizing across distinct fiber types due to feature distribution shifts. The proposed framework uniquely integrates the Vision Transformer's self-attention mechanism for holistic microstructure interpretation with a domain adaptation strategy to explicitly address inter-fiber variability. This approach enables direct mapping from raw micrographs to bond properties without relying on handcrafted features, providing a unified solution that surpasses the capabilities of existing techniques in both prediction accuracy and cross-fiber generalization.

To address the aforementioned challenges of cross-fiber type prediction and the shortcomings of existing methods in multi-scale feature fusion and cross-domain generalization, this paper proposes a framework for predicting the interfacial bond strength of natural fiber concrete that combines the Vision Transformer with a domain adaptation mechanism. Compared to traditional image analysis methods, classic CNN (Convolutional Neural Network) models, and conventional machine learning methods, ViT possesses powerful global context modeling capabilities. Lei Wang's team [20] studied the mapping strategy used in the image patch embedding process and explicitly solved the conversion problem from two-dimensional (2D) to one-dimensional (1D) representation. Mwamba Kasongo Dahouda's team [21] proposed a deep learning embedding technique for encoding classification features on classification datasets. Nhat-Duc Hoang's team [22] constructed and verified a data-driven ultimate bond strength estimation method. Experimental results show that this method shows higher accuracy, lower error fluctuation and stronger interpretability in cross-domain prediction tasks of multiple natural fiber types, achieving end-to-end modeling from microscopic

images to bonding properties.

ALGORITHM DESIGN

The proposed analysis framework's algorithm design comprises four core components: image preprocessing, network architecture, feature encoding, and generalization mechanism.

Image Acquisition and Preprocessing

In this study, microscopic images were acquired using a scanning electron microscope (SEM, JSM-IT800) and a high-resolution optical microscope (Nikon Eclipse LV100). 16-bit grayscale images were acquired at 500× and 1000× magnifications for each natural fiber-cement matrix interface. Images were taken at a uniform illumination intensity (3200 lx) and working distance (10 mm) to eliminate structural deviations introduced by varying imaging conditions.

After acquisition, the original image is first subjected to non-local means (NLM) filtering to denoise it. The core idea is to use the weighted average of similar blocks in the image to suppress noise. For image I, the denoising result of pixel p is given by the following formula:

$$\hat{I}(p) = \frac{\sum_{q\in\Omega} \omega(p,q) I(q)}{\sum_{q\in\Omega} \omega(p,q)}, \omega(p,q) = \exp\left(-\frac{\|I(N_p) - I(N_q)\|_2^2}{h^2}\right) \quad (1)$$

N_p and N_q are the neighborhood blocks centered on p and q, respectively. h is the filter strength parameter (set to 0.15 in this experiment), and Ω is the search window (radius 10 pixels).

For residual periodic interference patterns, the BM3D algorithm is used. Its hard thresholding step can be formalized as follows:

$$\widetilde{X} = T_\lambda\left(\mathrm{F}^{-1}\left[\mathrm{F}(G)\cdot \mathbf{1}_{|\mathrm{F}(G)|\geq\lambda}\right]\right) \quad (2)$$

G represents the grouped 3D image blocks, F and F^{-1} represent the 3D forward and inverse transform operators, respectively, and $T_\lambda(\cdot)$ represents the hard thresholding operator.

Subsequently, all images are uniformly scaled to 224×224 pixels using bicubic interpolation to ensure that the

input size matches the Vision Transformer's patch structure. For spatial continuity modeling, the bicubic interpolation kernel function can be expressed as:

$$k(x)=\begin{cases}(a+2)|x|^3-(a+3)|x|^2+1, & 0\le|x|<1\\ a|x|^3-5a|x|^2+8a|x|-4a, & 1\le|x|<2\\ 0, & |x|\ge2\end{cases} \tag{3}$$

The variable x represents the input to the interpolation kernel function. The parameter a controls the shape of the kernel function, and is set to -0.5 in this study.

To reduce sampling bias, each image is segmented into 196 non-overlapping patches (16×16 pixels) using a fixed-step partitioning method. Centering and standard deviation normalization are performed before segmentation:

$$I'_{i,j}=\frac{I_{i,j}-\mu}{\sigma+\epsilon} \tag{4}$$

$I'_{i,j}$ is the normalized pixel value, μ and σ are the grayscale mean and standard deviation of the current image, respectively. ϵ is a small constant to prevent the denominator from being zero. This ensures consistency in grayscale distribution across batches of images and improves the stability and generalization of the Vision Transformer during the feature encoding phase. Image sample statistics are shown in Table 1.

Table 1. Image Sample Statistics

Fiber Type	Magnification	Number of Samples	Image Resolution (bit)	Light Intensity (lx)	Working Distance (mm)
Jute Fiber	500×	180	16-bit	3200	10
Jute Fiber	1000×	180	16-bit	3200	10
Sisal Fiber	500×	175	16-bit	3200	10
Sisal Fiber	1000×	175	16-bit	3200	10
Coir Fiber	500×	160	16-bit	3200	10

Coir Fiber	1000×	160	16-bit	3200	10
Flax Fiber	500×	170	16-bit	3200	10
Flax Fiber	1000×	170	16-bit	3200	10
Hemp Fiber	500×	165	16-bit	3200	10
Hemp Fiber	1000×	165	16-bit	3200	10
Total	—	1700	—	—	—

The selection of non-local means filtering and BM3D denoising is driven by their proven ability to reduce noise while preserving edge integrity and textural details in microscopic images. Non-local means filtering leverages spatial redundancy to average similar patches, minimizing noise without smoothing critical interface morphology. BM3D targets residual periodic artifacts through collaborative thresholding in transform domains, preventing the loss of high-frequency microstructural information. The 16×16 pixel patch size is adopted to match the input requirements of Vision Transformer architectures, ensuring a balance between computational efficiency and the capture of local features. This dimension allows each patch to encompass sufficient contextual information for global attention mechanisms while resolving fine-scale details such as fiber surface roughness or pore distribution. These preprocessing steps are optimized to enhance signal-to-noise ratio without obscuring essential microstructural characteristics, as evidenced by the model's high predictive accuracy and interpretable attention maps focused on interfacial regions.

Vision Transformer Architecture

During model construction, the 196 preprocessed patches are sequentially flattened into vectors and passed through a linear projection matrix:

$$W_{\text{embed}} \in \mathrm{R}^{(16\times16)\times384} \tag{5}$$

Projected into a 384-dimensional feature space to compress redundant pixel information and unify feature scales. The specific mapping process is:

$$Z_0=XW_{\text{embed}}+P_{\text{pos}} \tag{6}$$

$X\in R^{196\times(16\times16)}$ represents the input matrix after patch flattening. $XW_{\text{embed}}\in R^{(196\times384)}$ denotes the embedded patch tokens. A learnable classification token ([CLS]) is prepended to this sequence, resulting in a token sequence of length 197. $P_{\text{pos}}\in R^{197\times384}$ is a learnable positional encoding matrix added to this sequence to preserve spatial structure information.

The encoder uses a 12-layer stacked Transformer block to fully model the long-range dependencies of the fiber-matrix interface. Each layer contains a multi-head self-attention (MHSA) module and a feed-forward network (FFN). The core calculation formula of the MHSA is:

$$\text{Attention}(Q,K,V)=\text{softmax}\left(\frac{QK^{\text{T}}}{\sqrt{d_k}}\right)V \tag{7}$$

The query matrix $Q=ZW_Q$, the key matrix $K=ZW_K$, and the value matrix $V=ZW_V$, respectively, have weight matrices $W_Q,W_K,W_V\in R^{384\times64}$. The dimension of a single head is 64, and $d_k=64$ is the scaling factor.

The multi-head mechanism uses 6 independent attention heads (i.e., head_1 to head_6) to capture interface structural features at different scales in parallel. The outputs are concatenated and then linearly transformed:

$$\text{MHSA}(Z)=\text{Concat}(\text{head}_1,\ldots,\text{head}_6)W_O,\quad \text{head}_i=\text{Attention}(Q_i,K_i,V_i) \tag{8}$$

where $W_O\in R^{(6\times64)\times384}$.

The feedforward network in the encoder is expressed as a two-layer fully connected structure, and the activation function uses GELU (Gaussian Error Linear Unit), specifically:

$$\text{FFN}(x)=\text{GELU}(xW_1+b_1)W_2+b_2 \tag{9}$$

$W_1\in R^{384\times1536}$, $W_2\in R^{1536\times384}$, b_1,b_2 are bias terms. The sequence of patch embeddings is extended by

prepending a learnable classification token, resulting in a total sequence length of 197. This structure enhances nonlinear mapping capabilities and improves the model's ability to discriminate complex interface textures. The model architecture is shown in Figure 1.

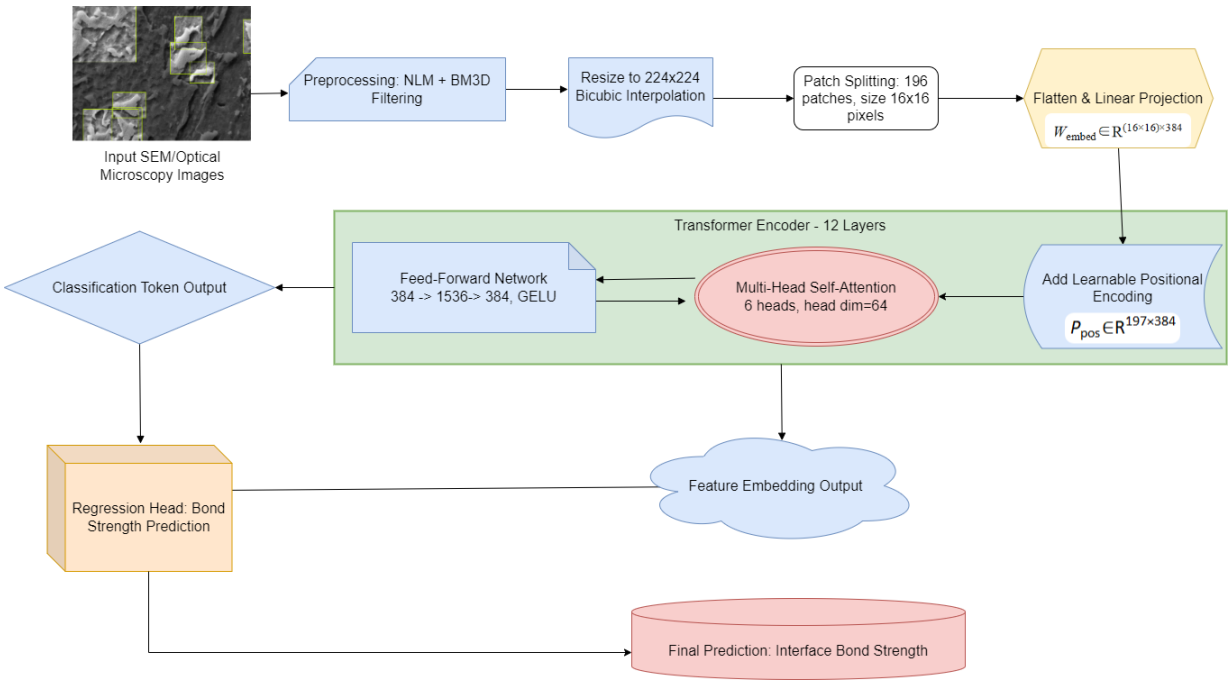

Figure 1. Model Architecture

The Vision Transformer model is initialized using weights pre-trained on the dataset. During training, the model parameters are fine-tuned end-to-end on the natural fiber concrete dataset. To mitigate overfitting due to the limited sample size, strong regularization strategies are employed, including Dropout, weight decay in the AdamW optimizer, and early stopping based on validation loss.

Feature Extraction and Encoding

The sequence after patch embedding and positional encoding is input to the Vision Transformer encoder. At each layer, a multi-head self-attention mechanism is used to separate and associate local interface textures with global adhesion patterns. To account for feature representation at different depths, the [CLS] token and the mean features of all patches are extracted from the output sequence at layers 4, 8, and 12, respectively. The output sequence of layer l is defined as:

$$Z^{(l)}=\{z_{\text{cls}}^{(l)},z_1^{(l)},z_2^{(l)},\ldots,z_{196}^{(l)}\},z_i^{(l)}\in \mathrm{R}^{384} \tag{10}$$

The mean feature vector is expressed as:

$$\bar{z}^{(l)}=\frac{1}{196}\sum_{i=1}^{196} z_i^{(l)} \tag{11}$$

After extraction, perform LayerNorm normalization on $z_{\text{cls}}^{(l)}$ and $\bar{z}^{(l)}$ respectively:

$$\hat{z}^{(l)}=\text{LayerNorm}(z_{\text{cls}}^{(l)}\oplus\bar{z}^{(l)}) \tag{12}$$

Here, $\oplus$ represents the concatenation operation, resulting in a vector of dimension 768.

Subsequently, channel compression is performed using a 1×1 convolution mapping function f_{conv}:$\mathrm{R}^{768}\rightarrow\mathrm{R}^{512}$, expressed as a linear transformation:

$$f_{\text{conv}}(\hat{z}^{(l)})=W_{\text{conv}}\hat{z}^{(l)}+b_{\text{conv}},W_{\text{conv}}\in \mathrm{R}^{512\times768} \tag{13}$$

Normalization of the three-layer features is performed using learnable weight vectors and Softmax.

This multi-scale fusion mechanism enhances the model's ability to represent interface microstructure by integrating semantic information at different levels. Low-level features focus on microscopic morphologies such as fiber surface micropores and local roughness, while high-level features model global structures such as interface crack propagation paths and bond failure modes. The weighted fusion of these two features achieves multi-scale feature synergy from local to global scales, providing a more comprehensive physical basis for bond strength prediction.

Finally, the fused features $\tilde{Z}$ are subjected to a global average pooling operation GAP(·) to produce a fixed-length global image representation $z_{\text{final}}\in\mathrm{R}^{512}$, which is used for subsequent regression prediction:

$$z_{\text{final}}=\text{GAP}(\tilde{Z}) \tag{14}$$

The feature extraction and encoding parameters are shown in Table 2.

Table 2. Feature extraction and encoding parameters

Layer Index	Feature Type	Dimension Before Concat	Dimension After Concat	After 1×1 Conv (512-d)	Learned Weight (α)	Normalized Weight (Softmax)
4	[CLS] Token + Mean	384 + 384	768	512	0.92	0.36
8	[CLS] Token + Mean	384 + 384	768	512	1.12	0.39
12	[CLS] Token + Mean	384 + 384	768	512	0.81	0.25

Design of a Multi-Fiber Source Generalization Mechanism

The adversarial domain adaptation structure is designed to align the feature distributions across different natural fiber types, each of which constitutes a distinct domain. The primary challenge is that data from one fiber type, such as jute, exhibits a different statistical distribution in feature space compared to another, such as sisal. This domain shift impedes model generalization. The adaptation mechanism explicitly addresses this by defining the domain discriminator to classify the fiber type origin of the input features. Simultaneously, the feature extraction backbone is trained to confuse this discriminator through a gradient reversal layer. This adversarial process compels the model to learn domain-invariant representations that are predictive of bond strength but indistinguishable with respect to the fiber type, thereby enhancing cross-fiber generalization. The implementation combines this with Domain-Specific Batch Normalization to handle differing feature statistics across domains.

At the model input stage, a learnable 128-dimensional embedding vector is assigned to each natural fiber and concatenated with each image patch sequence. After extracting the multi-scale features output by the feature extraction module, a learnable 128-dimensional embedding vector E_f is first assigned to each natural fiber. This is then concatenated with the global image feature vector F_g in the channel dimension to form a joint representation:

$$Z=[F_g;E_f],Z\in \mathrm{R}^{768+128} \tag{15}$$

Subsequently, the linear mapping matrix $W_z\in R^{(768+128)\times 768}$ is used to map back to the unified 768-dimensional feature space:

$$F_{joint}=ZW_z+b_z,F_{joint}\in \mathrm{R}^{768} \tag{16}$$

The combined feature F_{joint} is fed into the backbone network for adhesion strength regression prediction. This fused feature replaces the original patch features and is subsequently fed into the Vision Transformer encoder backbone network.

During training, a gradient reversal layer (GRL) is inserted before the domain discriminator. It is defined as:

$$\mathrm{GRL}(x)=x,\frac{\partial \mathrm{GRL}}{\partial x}=-\lambda I \tag{17}$$

This mechanism reverses the sign of the gradient during backpropagation, prompting the backbone network to learn domain-indistinguishable feature representations and achieve feature distribution alignment.

To mitigate differences in batch normalization statistics between different fibers, Domain-Specific Batch Normalization (DSBN) is employed. The input feature x is calculated domain by domain:

$$\hat{x}^{(d)}=\frac{x^{(d)}-\mu^{(d)}}{\sqrt{(\sigma^{(d)})^2+\epsilon}}\cdot\gamma^{(d)}+\beta^{(\mathrm{d})} \tag{18}$$

Here, d represents the fiber domain category, $\mu^{(d)}$ and $\sigma^{(d)}$ are the domain-independent mean and standard deviation parameters, $\gamma^{(d)}$ and $\beta^{(d)}$ are learnable scale and bias, and ϵ is a minimal constant to prevent division by zero.

The training objective is the weighted sum of the regression loss L_{reg} and the domain adversarial loss L_{adv}, with the weight λ being linearly increased:

$$L_{total}=L_{reg}+\lambda(t)L_{adv},\lambda(t)=0.5\cdot\frac{t}{T} \tag{19}$$

Here, t is the current training round number, and T is the total training round number. This ensures that the domain adversarial effect gradually increases, avoiding negative impacts on the convergence of the main task in the early stages. The generalization process is shown in Figure 2.

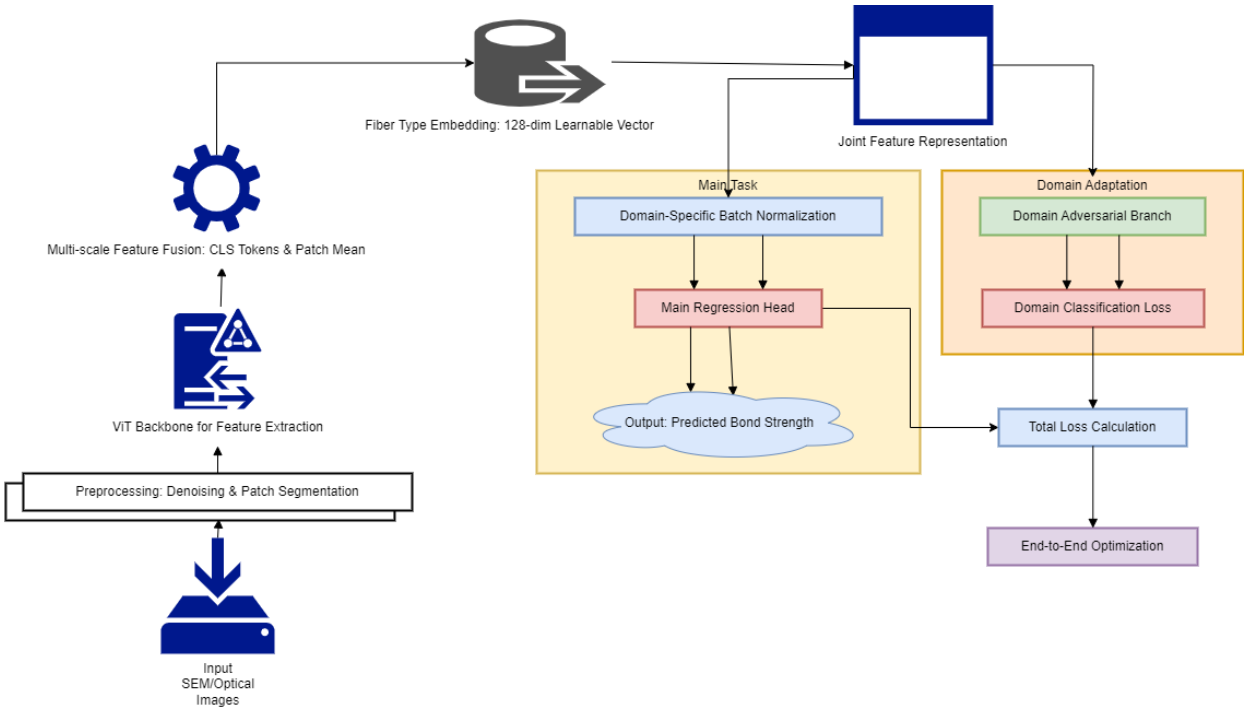

Figure 2. Generalization Process Schematic

Regression Head Construction and Training Strategy

The global feature vector $\mathbf{z}\in R^{768}$ output by the feature fusion module is first connected to a two-layer fully connected network to form a regression prediction head. The first layer has a 256-dimensional linear projection, which is activated by the GELU function and expressed as:

$$\mathbf{h}=\mathrm{GELU}(\mathbf{W}_1\mathbf{z}+\mathbf{b}_1),\mathbf{h}\in R^{256} \tag{20}$$

$\mathbf{W}_1 \in R^{256\times768}$, $\mathbf{b}_1 \in R^{256}$ are weight and bias parameters. To alleviate overfitting, a Dropout operation is applied with a probability of 0.1. The second layer maps the 256-dimensional features into a single scalar prediction value:

$$\hat{y}=\mathbf{W}_2\mathbf{h}+b_2,\hat{y}\in \mathrm{R} \tag{21}$$

where $\mathbf{W}_2 \in R^{1\times256}$, $b_2 \in R$.

The Huber loss function is used during training (taking into account the advantages of both MAE and MSE and adapting to the non-uniform distribution of bond strength). Its segmented form is defined as:

$$L_{\mathrm{Huber}}=\frac{1}{N}\sum_{i=1}^{N}\begin{cases}\frac{1}{2}(y_i-\hat{y}_i)^2, & \mathrm{if}|y_i-\hat{y}_i|\le\delta\\ \delta|y_i-\hat{y}_i|-\frac{1}{2}\delta^2, & \mathrm{otherwise}\end{cases} \tag{22}$$

The variable y represents the actual bond strength value, and ŷ represents the model prediction value. δ is the threshold parameter, set to 0.1. N is the batch size.

The optimizer uses the AdamW algorithm, with a weight decay coefficient of 0.01, an initial learning rate of 1×10^{-4}, and a cosine annealing scheduling strategy:

$$\eta_t = \eta_{\min} + \frac{1}{2}(\eta_{\max} - \eta_{\min})\left(1 + \cos\left(\frac{t\pi}{T}\right)\right) \tag{23}$$

η_t is the learning rate for iteration t, $\eta_{max}=1\times10^{-4}$, η_{min} are the minimum learning rates, and T is the total number of iterations. An early stopping mechanism is also introduced: training is terminated when the validation set MAE does not decrease by more than 0.001 for 10 consecutive epochs, and the optimal weights are restored. The training architecture is shown in Figure 3.

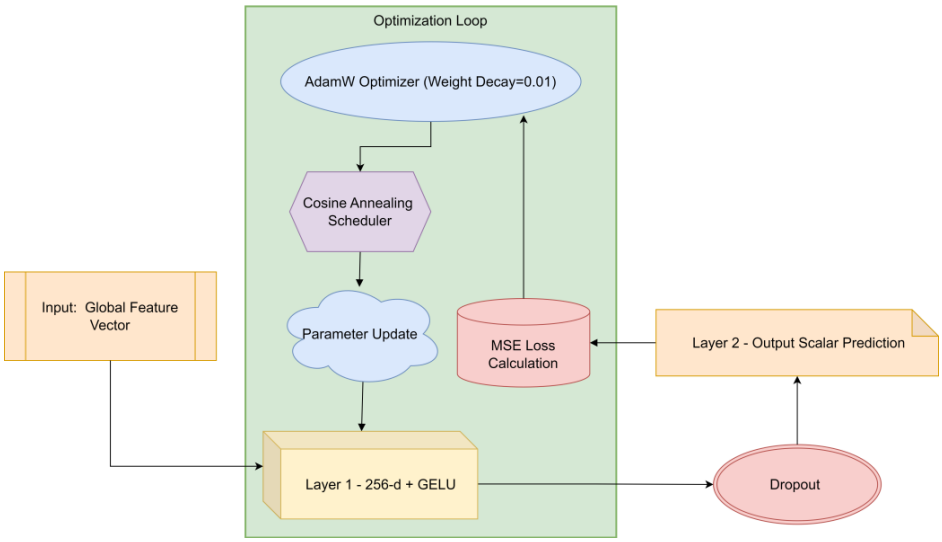

Figure 3. Training Architecture

EXPERIMENTS AND VALIDATION

Based on the aforementioned algorithm framework, this section evaluates the model's performance in terms of bond strength prediction, generalization across different fiber types, and feature interpretability through systematic experiments.

Correlation Analysis between Microscopic Image Features and Bond Strength

The experimental validation utilized three hundred concrete specimens, each incorporating a single natural fiber type. The sample set comprised one hundred hemp, eighty flax, sixty jute, forty coir, and twenty sisal fiber specimens. Following a standard curing period of twenty-eight days, each specimen underwent mechanical pull-out and splitting tests to determine the reference bond strength. All natural fibers were treated with a 5% sodium hydroxide solution for two hours. The cementitious matrix mix ratio was cement:water:standard sand = 1:0.5:1.5. Specimens were standard cubic blocks, 100 mm × 100 mm × 100 mm in size. Interfacial bond strength was determined through a splitting tensile test conducted in accordance with ASTM C496 standard. The test was conducted on a universal testing machine with a loading rate controlled at 0.5 MPa/s. For microscopic analysis, a section containing the fiber-matrix interface was extracted from each tested specimen. These sections were polished to an optical finish. From each sample section, multiple microscopic images were systematically captured along the fiber-matrix interface at predetermined

intervals to ensure comprehensive coverage and representativeness. This procedure guaranteed that the image data encompassed the variability inherent in the interfacial zone. The total collection yielded one thousand seven hundred images, with the distribution per fiber type and magnification detailed in Table 1. The direct correlation between each image set and the mechanically measured bond strength of its parent specimen forms the foundation for the subsequent predictive modeling.

In the experimental design, a correspondence between image samples and mechanical test results was established for different types of natural fiber interfacial bonding. All microscopic images were standardized and evenly divided by fiber type to ensure representative sample distribution during model training, validation, and testing. The overall sample distribution is shown in Table 3.

Table 3. Sample Category Statistics

Fiber Type	Total Samples	Training Set (70%)	Validation Set (15%)	Test Set (15%)	Bond Strength Label Source
Hemp	100	70	15	15	Pull-out & Splitting Test
Flax	80	56	12	12	Pull-out & Splitting Test
Jute	60	42	9	9	Pull-out & Splitting Test
Coir	40	28	6	6	Pull-out & Splitting Test
Sisal	20	14	3	3	Pull-out & Splitting Test
Total	300	210	45	45	—

Bond Strength Prediction Error Metrics

The high predictive accuracy observed stems from a rigorously designed data separation strategy and the model's intrinsic capability to capture deterministic interfacial features. The train-test split was strictly performed at the sample level, ensuring that all microscopic images originating from the same physical specimen were allocated exclusively to either training, validation, or test sets. This sample-wise partitioning prevents data leakage by guaranteeing the model is evaluated on entirely unseen specimens. The exceptional performance metrics are attributed to the Vision Transformer's effectiveness in learning the strong underlying

correlation between highly detailed interfacial morphology, as captured in high-resolution micrographs, and the resulting bond strength. The model succeeds in identifying these physically meaningful, deterministic patterns rather than memorizing spurious correlations, a capability enhanced by the domain adaptation mechanism which forces the learning of domain-invariant features. The minimal performance gap between training and test sets further supports the absence of overfitting and validates the model's generalization capacity.

After completing the test set predictions, the predicted values were paired with the measured bond strengths one by one. Five metrics, namely MAE (Mean Absolute Error), RMSE (Root Mean Square Error), R^2, MAPE (Mean Absolute Percentage Error), and RMSE% were calculated using batch operations.

MAE and RMSE are calculated by averaging the absolute and squared values of the prediction residuals. Residual calculations were performed using a vectorized approach on the GPU (Graphics Processing Unit) to ensure numerical accuracy and computational efficiency. $R^2$is computed as the coefficient of determination, measuring the proportion of variance in the measured bond strength that is explained by the model predictions. MAPE and RMSE% use a minimum lower bound of 0.001 on the denominator to avoid amplification caused by low-strength samples. All metrics are calculated on the same normalized scale, with bond strength values normalized to the range from zero to one based on the minimum and maximum values in the dataset, to eliminate the impact of different strength ranges on error quantification.

These metrics are calculated independently for each fiber type in the test set and a weighted average is taken globally to ensure that the evaluation results reflect both overall performance and balance across categories. Finally, the numerical values of each indicator were compared with the performance of the training set to determine the regression stability of the model on unseen data. The residual distribution was then combined to analyze the deviation trend of the model in different strength ranges, thereby identifying potential deficiencies in feature extraction and regression. The bond strength prediction performance indicators and residual distribution are shown in Figure 4.

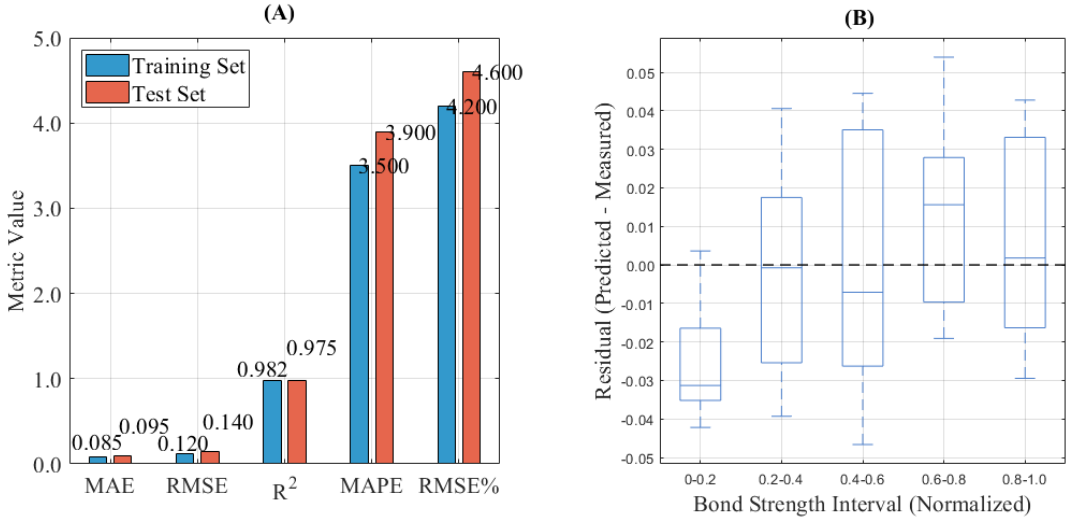

Figure 4. Bond Strength Prediction Performance Indicators and Residual Distributions: (A) Comparison of Model Prediction Performance Indicators; (B) Residual Distributions for Different Bond Strength Ranges

Figure 4A is a bar chart showing the comparison. The numerical values show that the MAE for the training and test sets are 0.085/0.095, the RMSE is 0.120/0.140, the R^2 is 0.982/0.975, the MAPE is 3.5%/3.9%, and the RMSE% is 4.2%/4.6%. Error metrics on the test set were only slightly higher than those on the training set, with improvements of less than 0.03 or 0.5 percentage points. R^2 remained above 0.97, demonstrating a small generalization gap and stable fitting capability. Figure 4B shows a boxplot of residuals for different normalized bond strength ranges, with the dashed line representing the zero residual baseline. The median is approximately −0.031 in the low strength range, −0.001 in the 0.2–0.4 range, and close to −0.007 in the 0.4–0.6 range. It becomes slightly positively skewed to approximately +0.016 in the 0.6–0.8 range, and drops back to approximately +0.002 in the 0.8–1.0 range.

Comparative Generalization Performance Metrics

To verify the model's predictive ability for unseen fiber types, the experiment employed a leave-one-out cross-validation (LOOCV) strategy.

In each round of experiments, all samples of a specific fiber type were completely removed from the training set, and only samples of that type were used for prediction during the test phase. The remaining data partitioning and hyperparameter settings remained unchanged during the training process to ensure comparability across experiments. After predictions were completed, the MAE, RMSE, and R^2 for that fiber

type were calculated and then subtracted from the average error of the remaining types to obtain the generalization error delta for that type. The standard deviation (Std) of the prediction results for all unseen types was also calculated to measure the model's stability in cross-fiber prediction. To reduce random fluctuations within a single partition, all experiments were repeated five times, and the average was taken as the final metric.

This process maintains feature distribution consistency through domain-adaptive feature alignment and category embedding, significantly reducing generalization error while also minimizing performance fluctuations across different fiber domains. A comparison of the model's generalization capabilities across different fiber types is shown in Figure 5.

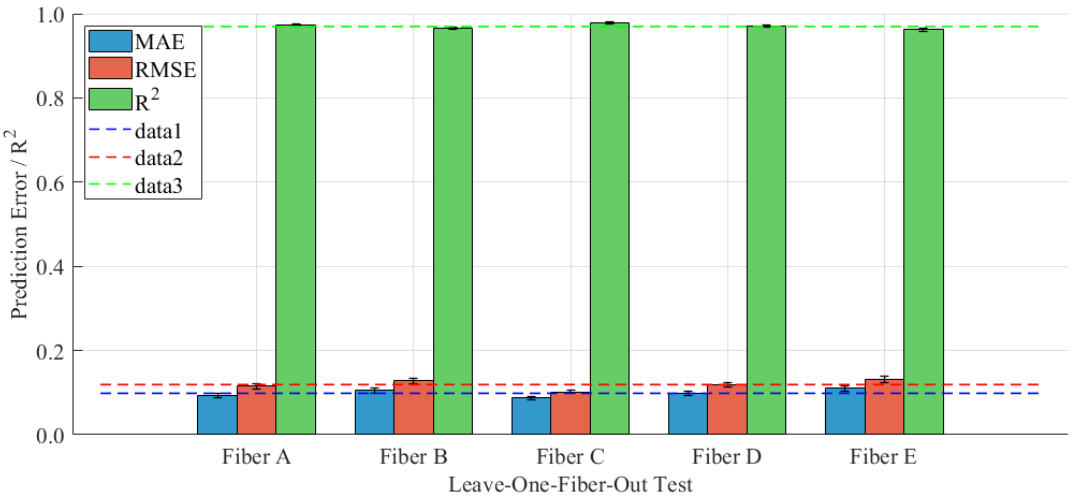

Figure 5. Comparison of Model Generalization Ability for Different Fiber Types

The horizontal axis of Figure 5 represents different natural fiber types (A to E), corresponding to hemp, flax, jute, coir, and sisal, respectively. The vertical axis represents the error and R^2. The data shows that Fiber C has the best generalization performance: it has the lowest MAE increment (0.087±0.004), an RMSE increment of only 0.102±0.005, and a high R^2 of 0.978±0.002, with a low standard deviation. In contrast, Fiber E performed poorly, with a MAE increase of 0.110±0.007 and an RMSE increase of 0.132±0.008. Fiber A (MAE 0.092±0.005) and Fiber D (0.098±0.005) were in the middle, while Fiber B (0.105±0.006) was near the lower limit of performance. The length of the error bars intuitively reflects the difference in stability. For example, the standard deviation of the MAE for Fiber E (0.007) is 1.75 times that of Fiber C (0.004), indicating that the model's predictions for the former are more volatile.

Feature Interpretability Evaluation Metrics

During the testing phase, the trained Vision Transformer model processed the test set images. To visualize the decision-making process, a Grad-CAM-style attribution method adapted for regression and Vision Transformers was employed. The gradient of the predicted bond strength with respect to the patch token embeddings from the final Transformer encoder layer was computed. These gradients were averaged across the token dimension to obtain importance weights for each patch token. The weighted token representations were then reshaped into a two-dimensional patch grid and passed through a Rectified Linear Unit to highlight regions with a positive contribution to the prediction. The resulting low-resolution heatmap was upsampled to 224×224 pixels using bilinear interpolation and superimposed onto the original microscopic images.

The heatmap was then interpolated to a 224×224 resolution and overlaid with the original microscopic image to generate a visual representation of the region of interest. To further pinpoint the model's spatial attention patterns, the weights of all attention heads are averaged along the channel dimension and mapped back to the original image patch coordinates using the positional encoding index to calculate the distribution of attention intensity within the interface region. Each image's heatmap is binarized (with a threshold set to 60% of the maximum value). The proportion of pixels in the hotspot region within the entire image is counted, and its mean and standard deviation are calculated on the test set to measure the concentration and stability of attention.

This process verifies that the model consistently focuses on the physically relevant fiber-matrix interface region across different fiber types and interface morphologies. It also quantifies the fluctuations in the pattern of interest across samples, thereby identifying potential sources of bias during the feature extraction phase. Figure 6 shows the visualization of fiber-matrix interface features and statistical analysis of the model's region of interest.

The heatmap in Figure 6A shows a hotspot surrounding the outer edge of the fiber and extending into the interface transition layer. A high-response core is also visible within the fiber, while a large area of the matrix is characterized by a green-blue, low-response region. Figure 6B shows the statistical results of model interest regions for three fiber types, with the horizontal axis representing fiber type and the vertical axis representing the proportion of hotspot regions. The data shows that coconut fiber has the lowest mean hotspot ratio (0.084 ± 0.025), with sample values concentrated between 0.05 and 0.13. Jute fiber has a higher mean of

0.116 ± 0.038, with a more dispersed distribution (0.05-0.1863). Sisal fiber has the highest mean (0.177 ± 0.055), with significant fluctuations.

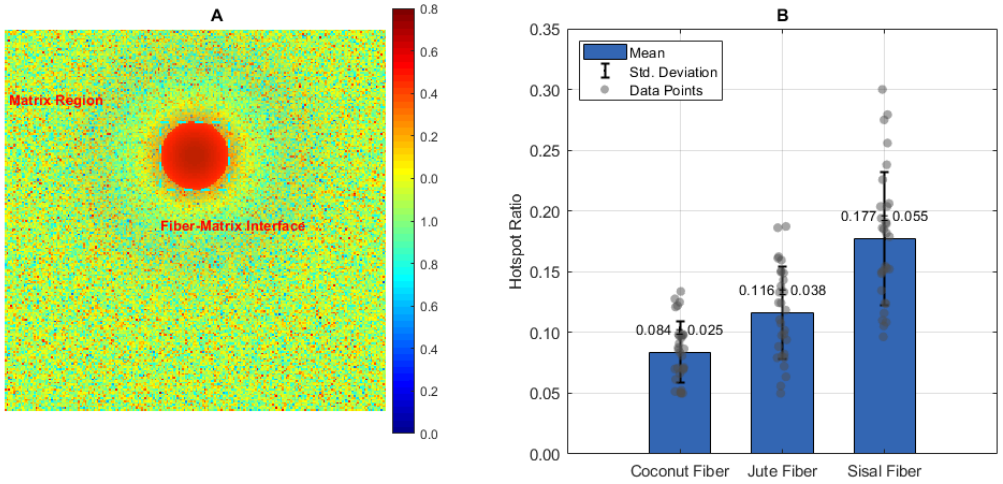

Figure 6. Visualization of Fiber-Matrix Interface Characteristics and Statistical Analysis of Model Interest Regions (ROIs): (A) Fiber-Matrix Interface Heatmap; (B) Statistics of ROIs Across Fiber Types

Comparison with Traditional Methods

To comprehensively evaluate the performance of the proposed ViT model, this experiment used the same training and test sets as the Vision Transformer model to compare bond strength predictions with various baseline models. The baseline models cover different paradigms: (1) Traditional image methods combined with machine learning: Texture parameters such as contrast, energy, entropy, and homogeneity are extracted based on the Gray-Level Co-occurrence Matrix (GLCM) of microscopic images. Feature vectors are then constructed based on the statistical geometric features of the interface morphology. These vectors are then input into a Support Vector Regression (SVR) model (with an RBF (Radial Basis Function) kernel, C=10, and γ=0.1) for prediction. (2) Classic deep learning models include ResNet50 (initialized using ImageNet pre-trained weights, replacing fully connected layers for regression) and Swin-Transformer (Tiny version, patch size 4×4, window size 7, number of layers {2, 2, 6, 2}); (3) The machine learning method retains the original random forest regression (RF), with 500 trees, a maximum depth of 20, and a minimum number of samples per node. All baseline models used the same input resolution of 224×224 pixels as the Vision Transformer. The ResNet50 and Swin-Transformer models were initialized using weights pre-trained on the dataset and fine-tuned end-to-end during training without freezing any layers. All models used the same training

configuration: a learning rate of 5×10^{-5}, a batch size of 32, and training epochs controlled by an early stopping mechanism that terminated training when the validation set loss did not decrease within 10 consecutive epochs. Data augmentation strategies included random horizontal flips and random rotations within ±10 degrees. These settings ensured fairness in model comparison.

After all models were trained, the three accuracy metrics of MAE, RMSE, and R^2 were calculated on the test set. The number of parameters for each model and the average inference time for a single microscopic image were also recorded to comprehensively evaluate the model's accuracy and engineering practicality. Performance evaluation was conducted on a workstation equipped with an NVIDIA Tesla V100 GPU, using the PyTorch framework with a computational precision of FP32. Inference time was measured using a batch size of 1. The results were compared item by item with the Vision Transformer model, and the relative reduction in MAE and RMSE, as well as the relative improvement in R^2, of ViT compared to each baseline model were calculated. To eliminate fluctuations caused by a single random split, all models were repeatedly trained and tested under five independent splits, and the average value was taken as the final comparison result to ensure fairness and reliability. The performance comparison of the natural fiber concrete interfacial bond strength prediction models is shown in Figure 7.

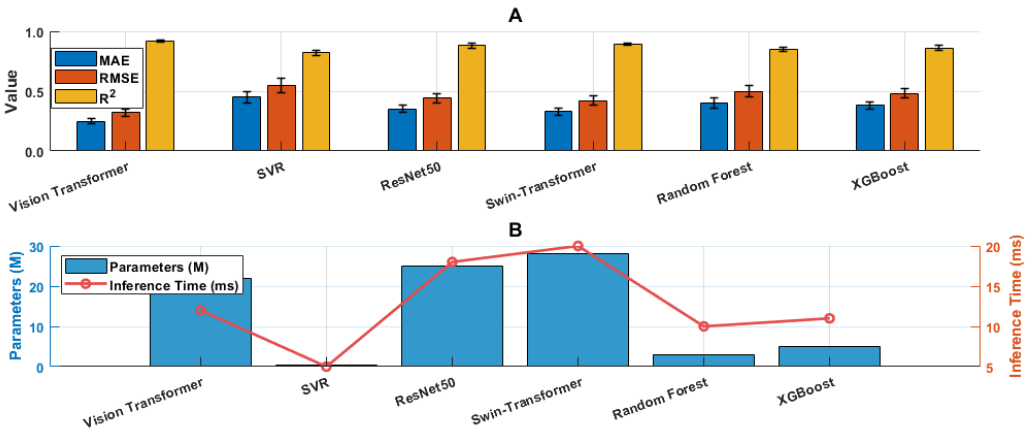

Figure 7. Performance Comparison of Natural Fiber Concrete Interface Bond Strength Prediction Models: (A) Comparison of Model Accuracy Indices (MAE/RMSE/R^2); (B) Comparison of Model Parameters and Inference Time

Sub-Figure A of Figure 7 shows the MAE, RMSE, and R^2 accuracy, with the model as the horizontal axis and the indicator value as the vertical axis. Sub-Figure B compares the number of parameters and single-image inference time. Compared to SVR's 0.45/0.55/0.82, MAE and RMSE decreased by 44.4% and 41.8%, respectively, while R^2 increased by 12.2%. Compared to ResNet50, the MAE decreased by 28.6%, the RMSE

decreased by 27.3%, and the R^2 increased by 4.5%. Compared to Swin-Transformer, the MAE decreased by 24.2%, the RMSE decreased by 23.8%, and the R^2 increased by 3.4%. Compared to Random Forest's 0.40/0.50/0.85 and XGBoost's 0.38/0.48/0.86, MAE decreased by 37.5%/34.2%, RMSE decreased by 36.0%/33.3%, and R^2 increased by 8.2%/7.0%. ViT has 22M parameters and an inference time of 12ms; ResNet50 and Swin-T (Swin Transformer) have inference times of 25M/18ms and 28M/20ms respectively; SVR is the lightest at only 0.5M and takes 5ms; Random Forest and XGBoost have inference times of 3M/10ms and 5M/11ms respectively.

CONCLUSION

The primary contribution of this study lies in the development of a Vision Transformer-based framework that offers a superior solution to the multiscale and cross-domain analysis of natural fiber concrete interfaces. It solves the critical problem of quantitatively linking microscopic interfacial features to macro-scale bond strength under conditions of significant material heterogeneity, a task where traditional methods exhibit inherent limitations. The framework's superiority stems from its ability to model global contextual relationships across the interface and its explicit design for domain-invariant feature learning, yielding a level of accuracy and generalization not achievable by existing image analysis or machine learning techniques. This paper addresses the cross-domain prediction challenge of interfacial bond properties of multi-source natural fiber concrete by constructing an end-to-end analysis framework based on ViT. Through block embedding and position encoding of microscopic images, the model effectively captures the long-range dependency characteristics of the fiber-matrix interface. This paper introduces fiber category embedding vectors to enhance feature recognition across fiber types. Combined with adversarial domain adaptation (GRL and DSBN), feature distribution alignment is achieved, improving the model's generalization. Experimental results show that the model achieves MAE=0.095, RMSE=0.140, and R^2=0.975 on the test set. In leave-one-class cross-validation, the generalization error remains low for unseen fiber types (MAE increment of 0.087±0.004, RMSE increment of 0.102±0.005, and R^2=0.978±0.002), verifying its robustness across fiber types. Grad-CAM visualization results further confirm that the model's focus areas are highly consistent with physically relevant interface characteristics. The hotspot ratio for coconut coir fiber is 0.084 ± 0.025, demonstrating the interpretability of the method. This research provides a reliable intelligent analysis tool for mix optimization

and interface modification in natural fiber concrete.

Author Contributions

Conceptualization – Chunmei Yao; methodology – Chunmei Yao and Xiangru Dong; investigation – Chunmei Yao and Xiangru Dong; writing-original draft preparation – Chunmei Yao and Xiangru Dong. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

Funding

This research has received funding from the following projects.

(1) Provincial Demonstration Virtual Simulation Training Base for Intelligent Construction, No. 2024sfxf009.

(2) Anhui Province First-class Core Golden Course - Building Construction Technology, No. 2024yljk019

Acknowledgements

Not applicable.

REFERENCES

[1] Farouk AIB, Jinsong Z. Prediction of interface bond strength between ultra-high-performance concrete (UHPC) and normal strength concrete (NSC) using a machine learning approach. Arabian Journal for Science and Engineering. 2022; 47(4):5337-5363. doi: 10.1007/s13369-021-06433-6

[2] Song H, Liu T, Gauvin F. Enhancing mechanical performance of green fiber cement composites: Role of eco-friendly alkyl ketene dimer on surfaces of hemp fibers. Journal of Materials Research and Technology. 2024; 28:3121-3132. doi: 10.1016/j.jmrt.2023.12.255

[3] Feng Y, Wang W, Wang S, Niu Z, Li L. Multi-scale analysis of mechanical properties of KH-560 coupling agent modified PVA fiber-rubber concrete. Composite Interfaces. 2023; 30(9):983-1010. doi: 10.1080/09276440.2023.2179267

[4] John VJ, Dharmar B. Influence of basalt fibers on the mechanical behavior of concrete—A review.

Structural Concrete. 2021; 22(1):491-502. doi: 10.1002/suco.201900086

[5] Feng Y, Feng J, Chen W, Zhao C, Li Z. Multi-scale analysis of styrene butadiene latex modified PVA fiber concrete. Journal of Thermoplastic Composite Materials. 2024; 37(9):3058-3083. doi: 10.1177/08927057231222285

[6] Gholampour A, Ozbakkaloglu T. A review of natural fiber composites: Properties, modification and processing techniques, characterization, applications. Journal of Materials Science. 2020; 55(3):829-892. doi: 10.1007/s10853-019-03990-y

[7] Lotfi A, Li H, Dao DV, Prusty G. Natural fiber–reinforced composites: A review on material, manufacturing, and machinability. Journal of Thermoplastic Composite Materials. 2021; 34(2):238-284. doi: 10.1177/0892705719844546

[8] Mohamad A, Qasim Z, Zhaoye Q, Babak S, Mohammed A. Finite element analysis of natural fibers composites: A review. Nanotechnology Reviews. 2020; 9(1):853-875. doi: 10.1515/ntrev-2020-0069

[9] Saad M, Agwa IS, Abdelsalam B, Amin M. Improving the brittle behavior of high strength concrete using banana and palm leaf sheath fibers. Mechanics of Advanced Materials and Structures. 2022; 29(4):564-573. doi: 10.1080/15376494.2020.1780352

[10] Ramu S, Gebremicheal GH, Mohan R, Kumar RS, Karthigairajan M, Masannan V, et al. Alkali and non-alkali treated coconut coir fiber-reinforced coconut shell powder/MWCNT-filled polyester matrix composite: An experimental comparison. Journal of Environmental Nanotechnology. 2024; 13(3):297-304. doi: 10.13074/jent.2024.09.243893

[11] Rocky BP, Thompson AJ. Analyses of the chemical compositions and structures of four bamboo species and their natural fibers by infrared, laser, and X-ray spectroscopies. Fibers and Polymers. 2021; 22(4):916-927. doi: 10.1007/s12221-021-0303-8

[12] He X, Zhou R, Liu Z, Yang S, Chen K, Li L. Review of research progress and development trend of digital image correlation. Multidiscipline Modeling in Materials and Structures. 2024; 20(1):81-114. doi: 10.1108/MMMS-07-2023-0242

[13] Subramanyam SP, Kotikula DK, Bennehalli B, Babbar A, Alamri S, Duhduh AA, et al. Plain-woven areca sheath fiber-reinforced epoxy composites: The influence of the fiber fraction on physical and mechanical features and responses of the tribo system and machine learning modeling. ACS Omega. 2024;

9(7):8019-8036. doi: 10.1021/acsomega.3c08164

[14] Tipu RK, Arora R, Kumar K. Machine learning-based prediction of concrete strength properties with coconut shell as partial aggregate replacement: A sustainable approach in construction engineering. Asian Journal of Civil Engineering. 2024; 25(3):2979-2992. doi: 10.1007/s42107-023-00957-y

[15] Xu H, Ma S, Hou Y, Zhang Q, Wang R, Luo Y, et al. Machine learning-assisted identification of copolymer microstructures based on microscopic images. ACS Applied Materials & Interfaces. 2022; 14(41):47157-47166. doi: 10.1021/acsami.2c15311

[16] Kohan L, Fioroni CA, Azevedo AGDS, Leonardi B, Baruque-Ramos J, Fangueiro R, et al. Jute textiles with enhanced interfacial bonding as reinforcement for cementitious composites. Journal of Composite Materials. 2024; 58(16):1847-1862. doi: 10.1177/00219983241249237

[17] Gudayu AD, Getahun DE, Mekuriaw DM, Walelign FT, Ahmed AS. Natural fiber reinforced cementitious composites; materials, compatibility issues and future perspectives. Composite Interfaces. 2025; 32(3):363-397. doi: 10.1080/09276440.2024.2417162

[18] Ahmad SA, Ahmed HU, Rafiq SK, Mohammed BK. Smart predictive modeling for compressive strength in sisal-fiber-reinforced-concrete composites: Harnessing SVM, GP, and ANN techniques. Multiscale Science and Engineering. 2024; 6(1):95-111. doi: 10.1007/s42493-024-00110-0

[19] Berrocal CG, Fernandez I, Rempling R. Crack monitoring in reinforced concrete beams by distributed optical fiber sensors. Structure and Infrastructure Engineering. 2021; 17(1):124-139. doi: 10.1080/15376494.2020.1731558

[20] Wang L, Tang XS, Hao K. GFPE-ViT: Vision transformer with geometric-fractal-based position encoding. The Visual Computer. 2025; 41(2):1021-1036. doi: 10.1007/s00371-024-03381-8

[21] Dahouda MK, Joe I. A deep-learned embedding technique for categorical features encoding. IEEE Access. 2021; 9:114381-114391. doi: 10.1109/ACCESS.2021.3104357

[22] Hoang ND, Tran XL, Nguyen H. Predicting ultimate bond strength of corroded reinforcement and surrounding concrete using a metaheuristic optimized least squares support vector regression model. Neural Computing and Applications. 2020; 32(11):7289-7309. doi: 10.1007/s00521-019-04258-x