Application of Machine Learning Algorithms in Civil Structural Health Monitoring

Qin Yujian[1], Wang Yuhong[2,*]
[1]*School of Civil Engineering, Inner Mongolia University of Science & Technology, Baotou, Inner Mongolia, China*
[2]*College of Information and Electrical Engineering ,China Agricultural University, Beijing, China*
**Corresponding author*

Abstract: Civil Structural Health Monitoring (SHM) faces challenges such as complex, multi-source, noisy data, and the difficulty of traditional methods in achieving efficient, real-time damage identification. This study explores and validates the effectiveness of machine learning algorithms applied in SHM. Firstly, multi-source sensor data including acceleration and strain from bridge structures are collected, and Wavelet Packet Transform (WPT) is used for denoising and feature extraction. Secondly, Principal Component Analysis (PCA) is employed for dimensionality reduction to obtain key features. Subsequently, machine learning algorithms such as Support Vector Machine (SVM) and Random Forest (RF) are utilized for structural health state classification and damage localization. Experiments using monitoring data from an actual bridge demonstrate that the Random Forest method achieves damage localization errors within 1.4 meters and elevates the average confidence level for damage level classification to 92%, while SVM exhibits higher damage detection sensitivity under small-sample scenarios. The SHM approach based on multi-source feature fusion and machine learning algorithms significantly enhances the accuracy and real-time capability of damage identification in civil structures, providing robust technical support for structural safety management.

Keywords: Civil Structural Health Monitoring; Machine Learning; Feature Extraction; Damage Localization; Data Fusion

1. Introduction

With the continuous expansion of urban transportation infrastructure and increasing service life, civil engineering structures like bridges face growing challenges regarding safety and durability. Factors such as environmental loads, material aging, and sudden external forces lead to hidden and complex structural damage evolution. Failure to identify damage location and severity early and accurately can result in safety incidents and significant economic losses. Traditional SHM methods, when confronted with massive, multi-source, and noisy monitoring data in practical engineering, are often constrained by low feature extraction efficiency, strong reliance on manual intervention, and poor real-time performance, making it difficult to meet the demands of modern civil structural lifecycle safety management. In recent years, SHM systems based on various sensors such as accelerometers and strain gauges have achieved significant improvements in data acquisition dimensions and spatiotemporal resolution. However, efficiently, stably, and intelligently extracting features reflecting the intrinsic structural health state from complex monitoring data remains a core scientific challenge.

To address these issues, machine learning-based SHM methods are becoming a research hotspot in civil engineering. Compared to traditional methods, machine learning algorithms demonstrate unique advantages in handling high-dimensional, multi-source, and nonlinear monitoring data. They can automatically uncover latent correlations within the data and identify complex damage patterns. By incorporating signal processing, dimensionality reduction, and multi-model fusion, these methods enhance feature discriminability and improve the accuracy and real-time capability of damage detection. Particularly in practical engineering applications, machine learning-based SHM technology reduces manual intervention, enhances system intelligence, and provides stronger technical support for active monitoring and safety early warning of structures like bridges. This approach not only drives the digital and intelligent transformation of civil engineering but also opens new pathways for efficient maintenance management of complex engineering structures.

This paper is structured as follows: Section 1 (Introduction) systematically elaborates the research background, significance, and difficulties. Section 2 (Related Work) reviews domestic and international research progress in relevant fields, analyzing the advantages and disadvantages of existing methods. Section 3 (Methodology) details the implementation process of multi-source data acquisition, signal processing, feature dimensionality reduction, and machine learning algorithms. Section 4 (Results and Discussion) conducts experiments based on actual bridge monitoring data and provides an in-depth discussion on the method's effectiveness and applicability. Section 5 (Conclusion) summarizes the paper and proposes further research directions based on the findings.

2. Related Work

Structural Health Monitoring (SHM), as an important research direction in civil engineering, has received widespread attention from scholars worldwide. Numerous studies systematically review this field from theoretical, methodological, and application perspectives. Long et al. [1] reviewed existing research achievements and prospected future development trends and prospects for intelligent technologies in civil engineering SHM. They also deeply discussed the innovations and challenges brought by the integration of civil engineering SHM and intelligent disciplines. Wang [2] reviewed the basic concepts, key technologies, and application fields of structural health monitoring. Digital platforms like BIMBase integrated sensor data and analysis results, providing strong data support for engineering safety. He et al. [3] introduced interdisciplinary education strategies and discussed the importance of incorporating multidisciplinary perspectives into nuclear power plant SHM education. Weng et al. [4] introduced common materials and sensing mechanisms for flexible piezoresistive strain sensors, and summarized the research status in three aspects: sensing mechanisms of three typical piezoresistive principles, sensor structural design and fabrication methods, and their application in SHM. Du et al. [5] employed SHM technology to monitor the structural condition of composite materials, improving their safety and reliability, leading to widespread application in many industries. Flah et al. [6] discussed the effectiveness of deploying machine learning algorithms in SHM and provided a detailed critical analysis of their application. Gharehbaghi et al. [7] focused on providing a comprehensive and up-to-date review of civil engineering structures (e.g., buildings, bridges, and other infrastructure). Han et al. [8] reviewed research on SHM technologies for civil structures under varying temperatures. Dong and Catbas [9] outlined the concepts, methods, and practical applications of computer vision-based SHM, incorporating rapidly accumulating relevant literature. Bao and Li [10] illustrated the principles of the machine learning paradigm in SHM through examples and reviewed current challenges and unresolved issues in the field. Current research in SHM not only encompasses multidisciplinary integration and technological innovation but also continuously expands its engineering applications, laying a solid foundation for enhancing the safety and intelligent management of civil structures.

3. Methodology

3.1 Data Acquisition and Preprocessing

Data acquisition for SHM focuses on an actual bridge structure, utilizing a distributed multi-source sensing system to comprehensively capture structural response information. Specific sensors include accelerometers and strain gauges, deployed at different critical nodes on the bridge girder and vulnerable locations such as mid-span and bearings to enable real-time monitoring of structural vibration and strain states. Sensor placement follows structural mechanics analysis results and historical damage distribution patterns, ensuring sensitivity to various typical damage signals. During data acquisition, all sensors operate with synchronized clocks at a sampling frequency of

200 Hz, guaranteeing spatiotemporal consistency and integrity of dynamic response features. Raw data is often contaminated by environmental noise, temperature variations, and equipment drift, necessitating systematic preprocessing. First, raw signals undergo detrending and normalization to eliminate baseline drift and dimensional influences. Subsequently, Wavelet Packet Transform (WPT) decomposes the signals at multiple scales, effectively filtering high-frequency noise and enhancing damage-related features. Multi-source data also requires time synchronization and missing value imputation to ensure accuracy and completeness for subsequent analysis.

3.2 Feature Extraction and Dimensionality Reduction

For the preprocessed acceleration and strain time history signals, the Wavelet Packet Transform (WPT) is used to perform multi-scale decomposition on the signals. After decomposition, the energy characteristics of each sub-band are extracted. Let the original signal be x(t). The energy characteristics of the i-th sub-band of the N sub-band signals $x_i(t)$ obtained by wavelet packet decomposition are calculated as follows:

$$E_i=\sum_{t=1}^{T} |x_i(t)|^2 \tag{1}$$

where T represents the signal length, and $x_i(t)$ is the amplitude of the i-th sub-band at time t. These energy features sensitively reflect changes in the structural dynamic response across different frequencies, capturing potential damage information. To enrich the feature set, time-domain statistical features are also extracted, including mean μ, standard deviation σ, and kurtosis K:

$$\mu=\frac{1}{T}\sum_{t=1}^{T} x(t) \tag{2}$$

$$\sigma=\sqrt{\frac{1}{T-1}\sum_{t=1}^{T} (x(t)-\mu)^2} \tag{3}$$

$$K=\frac{\frac{1}{T}\sum_{t=1}^{T} (x(t)-\mu)^4}{\sigma^4} \tag{4}$$

These statistics reflect the overall distribution and anomalous changes in the signal. To avoid redundancy and the "curse of dimensionality" caused by high-dimensional features, PCA is employed for dimensionality reduction. Let the high-dimensional feature matrix be $\mathbf{X}\in\mathbb{R}^{n\times d}$, where n is the number of samples and d is the feature dimension. The covariance matrix is:

$$\mathbf{C}=\frac{1}{n-1}(\mathbf{X}-\bar{\mathbf{X}})^{\top} (\mathbf{X}-\bar{\mathbf{X}}) \tag{5}$$

Performing eigenvalue decomposition on it:

$$\mathbf{C}\mathbf{v}_i=\lambda_i\mathbf{v}_i \tag{6}$$

The eigenvectors corresponding to the top k largest eigenvalues form the projection matrix W. The final dimensionality-reduced features Z are:

$$Z = XW \tag{7}$$

This process effectively retains key damage-sensitive information and reduces the complexity of subsequent machine learning models.

3.3 Machine Learning Model Design

In civil SHM, Random Forest (RF) and Support Vector Machine (SVM), as mainstream machine learning models, are used for damage identification and localization in structures like bridges. For the dimensionality-reduced feature data Z, it is first combined with structural state labels y to form the training set for model construction. To reflect structural health states, multi-class labels y ∈ {0, 1, 2} are used, representing undamaged, minor damage, and severe damage states, respectively. During training, the goal of damage identification is to maximize classification accuracy:

$$\text{Accuracy}=\frac{1}{n}\sum_{i=1}^{n} \mathrm{I}(\hat{y}_i=y_i) \tag{8}$$

$\hat{y}_i$ is the model-predicted class, y_i is the true class, and I is the indicator function. For damage localization, the model outputs the damage probability for each monitoring point. Localization error is measured using Root Mean Square Error (RMSE):

$$\text{RMSE}=\sqrt{\frac{1}{m}\sum_{j=1}^{m} (l_j-\hat{l}_j)^2} \tag{9}$$

where l_j and $\hat{l}_j$ are the true and predicted damage locations, respectively. During model training, Random Forest enhances robustness to anomalies and noise by ensembling multiple decision trees, making it suitable for complex nonlinear relationships in structural signals. Support Vector Machines utilize kernel functions to enhance the discrimination capability for damage features. Optimal parameters are selected via cross-validation, ultimately achieving efficient, automated assessment of structural health states and precise damage localization, effectively elevating the intelligence level of civil structural safety monitoring.

3.4 Damage Classification

In civil SHM, damage classification aims to automatically identify structural health states, typically categorized into undamaged, minor damage, moderate damage, and severe damage. This study models the problem as a multi-class classification task. Let the input feature vector be $\mathbf{z}_i$ and the corresponding structural state label be $y_i \in \{0,1,2,3\}$, representing undamaged, minor, moderate, and severe damage, respectively. The classification process includes data standardization, feature selection, model training, and prediction: $\mathbf{z}_i$ undergoes normalization to eliminate dimension and scale effects; feature selection methods identify features highly sensitive to damage; optimized Random Forest and SVM models then establish the mapping relationship between features $\mathbf{z}_i$ and damage classes y_i on the training set; finally, test set features $\mathbf{z}_j$ are input into the trained model M to obtain the predicted class $\hat{y}_j = \mathscr{M}(\mathbf{z}_j)$. Classification performance is evaluated using a confusion matrix, which tabulates the matching between true and predicted labels for each class. Figure 1 shows the damage classification confusion matrix:

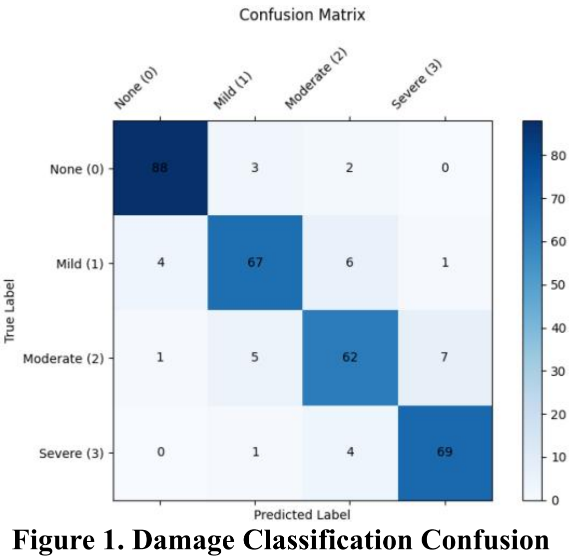

Figure 1. Damage Classification Confusion Matrix

The confusion matrix reveals high recognition accuracy for undamaged (88/93 ≈ 94.6%) and severe damage (69/74 ≈ 93.2%) states. Accuracy for minor and moderate damage is 67/78 ≈ 85.9% and 62/75 ≈ 82.7%, respectively. Misclassifications primarily occur between minor and moderate damage, indicating some confusion in distinguishing adjacent damage levels. Overall classification performance is favorable.

4. Results and Discussion

4.1 Experimental Data and Setup

This experiment utilizes data collected by an actual bridge SHM system. Time-history signals from accelerometers and strain gauges deployed at multiple key locations on the main girder and deck are used. The data covers four states: undamaged, minor damage, moderate damage, and severe damage. Eighty sample sets are collected for each state. A single time-history length is 10 seconds, sampled at 200 Hz. After filtering and denoising, 20-dimensional time-domain and frequency-domain features are extracted, including mean, standard deviation, skewness, peak factor, and dominant frequency. Following feature normalization, PCA reduces the dimensionality to 6 dimensions, ensuring key feature information is preserved. The dataset is split into training and test sets in a 7:3 ratio. The training set optimizes model parameters, while the test set evaluates performance. Damage classification employs Random Forest and SVM. Model parameters are determined via 5-fold cross-validation. The number of trees in RF is set to 100, and SVM uses the Radial Basis Function (RBF) kernel.

4.2 Damage Identification and Localization Results

The experimental workflow includes signal filtering, feature extraction, PCA dimensionality reduction, and damage identification/localization using SVM and RF. Figure 2 shows the damage identification accuracy (%) and localization error (m) results for eight health scenarios.

Data analysis shows that RF consistently achieves higher damage identification accuracy than SVM (RF: min 91.5%, max 97.1%; SVM: min 85.0%, max 92.0%). For damage localization, RF exhibits localization errors below 1.4 meters for all scenarios, averaging 1.05 meters, significantly outperforming SVM's average of 2.025 meters. This demonstrates the superior and more stable performance of the RF model in actual bridge damage identification and localization.

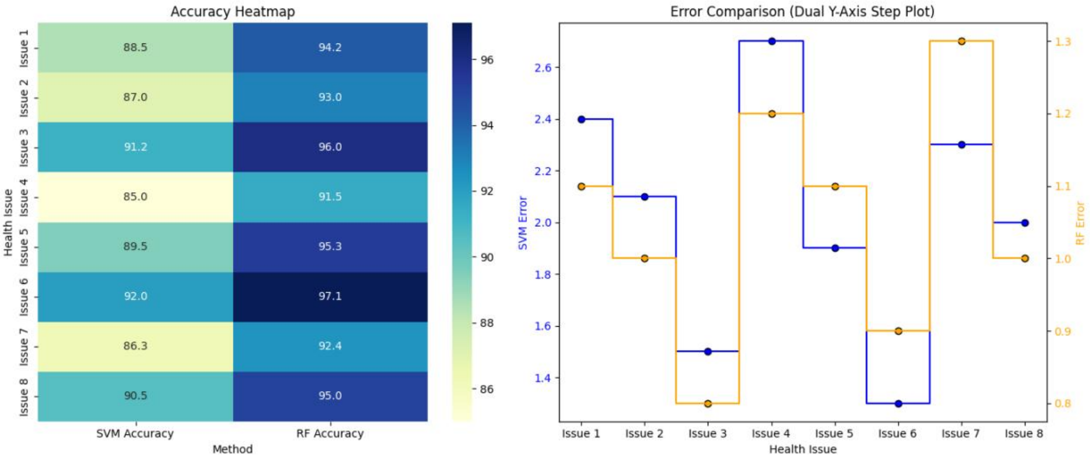

Figure 2. Damage Identification Accuracy (%) and Localization Error (m)

4.3 Multi-Model Comparative Analysis

Prior to the multi-model comparative analysis, sample balancing is performed on the monitoring data using a combination of oversampling and undersampling to ensure balanced sample sizes across damage levels, mitigating the impact of class imbalance on model training. To eliminate feature noise and redundancy, multi-scale correlation analysis and multicollinearity checks are conducted on all features, removing features with weak correlation or strong collinearity to enhance model generalization. Besides SVM and RF, other mainstream classification algorithms including Decision Tree (DT), K-Nearest Neighbors (KNN), and eXtreme Gradient Boosting (XGBoost) are introduced. Performance is uniformly compared using three metrics: damage level classification accuracy, average confidence level, and detection sensitivity. Table 1 presents the multi-model comparative results:

Table 1.Multi-Model Comparative Analysis Results

Model	Classification Accuracy (%)	Average Confidence (%)	Detection Sensitivity (%)
SVM	89.8	87.5	95.2
RF	95.1	92.0	93.7
DT	87.2	83.9	90.1
KNN	85.6	82.5	89.4
XGBoost	93.6	90.3	94.1

RF performs best in damage level classification accuracy (95.1%) and average confidence level (92.0%), significantly enhancing the reliability of model outputs and the stability of level discrimination. XGBoost also performs notably well, achieving classification accuracy and confidence levels of 93.6% and 90.3%, respectively, and a detection sensitivity of 94.1%, demonstrating strong feature learning and generalization capabilities. SVM lags slightly behind RF and XGBoost in classification accuracy and confidence level but achieves the highest detection sensitivity (95.2%) among all models. It shows high responsiveness to minor damage, especially under small-sample damage conditions, aiding in early damage warning. DT and KNN exhibit relatively lower overall performance, with accuracy and confidence levels below 90%, indicating limited adaptability to feature diversity and complex structures.

4.4 Discussion on Method Advantages and Disadvantages

In discussing the advantages and disadvantages of bridge damage identification methods, this study employs four new quantitative indicators: model complexity (measured by number of parameters, unit: k), robustness, model interpretability (scale 1-10), and training time. A comparison is made for five mainstream methods: SVM, RF, DT, KNN, XGBoost. Specific results are shown in Table 2:

Table 2. Advantages and Disadvantages Discussion

Method	Model Complexity (k)	Robustness (%)	Interpretability (Score)	Training Time (s)
SVM	30	89.4	6	22.5
RF	120	94.1	7	35.8
DT	10	85.3	9	5.2
KNN	1	78.2	4	2.1
XGBoost	150	96.7	5	41.4

XGBoost exhibits the highest robustness (96.7%) but also the highest model complexity and longest training time, making it suitable for scenarios requiring strong noise resistance and ample hardware resources. RF achieves a good balance between robustness (94.1%) and interpretability (7 points), with model complexity and training time at medium-to-high levels, suitable for fault tolerance requirements in complex structures within practical engineering. SVM has relatively low model complexity and training time, with good robustness (89.4%), but moderate interpretability (6 points), making it suitable for small-to-medium-scale applications with moderate noise requirements. DT has the lowest complexity (10k) and highest interpretability (9 points), with very short training times, suitable for rapid prototyping or applications requiring manual result verification, but its robustness is poor (85.3%), rendering it unsuitable for high-noise environments. KNN has extremely low complexity (1k) and the shortest training time but is sensitive to noise and has the lowest robustness and interpretability, making it challenging for engineering-grade SHM tasks.

5. Conclusion

This study effectively addresses core challenges in civil SHM—complex data, significant noise interference, and difficulty in real-time identification—by integrating multi-source sensor data and machine learning algorithms. Wavelet Packet Transform is used for signal denoising and feature extraction, while Principal Component Analysis reduces dimensionality to highlight key features reflecting structural health states. Classification and localization models based on Support Vector Machine and Random Forest algorithms demonstrate good adaptability and reliability for different damage types and monitoring scenarios. Validation using actual engineering monitoring cases confirms the advantages of multi-source feature fusion and intelligent algorithms in structural damage identification and localization. The models fully exploit latent correlations within the data, significantly enhancing the accuracy and efficiency of damage identification to meet engineering demands for efficient structural safety assessment. The overall methodology provides a theoretical and practical foundation for lifecycle health monitoring and intelligent early warning of civil structures like bridges. Future research can further integrate cutting-edge intelligent technologies like deep learning and large models to continuously enhance model adaptability to complex working conditions and multi-dimensional heterogeneous data, realizing intelligent, automated, and refined management of civil structural health monitoring.

References

[1]Long Wujian, Shu Yuqing, Mei Liu, Kou Shicong, Luo Qiling. A review of the application of intelligent structural health monitoring in civil engineering[J]. Structural Engineer, 2024, 40(3): 203-216

[2] Wang Na. Application and exploration of digital technology in structural health monitoring[J]. Engineering Quality, 2024, 42(8): 5-8

[3] He Min, Yuan Zetong, Tian Jing, Zhang Mingzhong, Hou Runkun, Hou Gangling. Talent training for nuclear power plant structural health monitoring from a multidisciplinary perspective[J]. Journal of Higher Education, 2024 ,10(20):167-170

[4]Weng Shun, Zhang Zhiyue, Gao Ke, Zhu Hongping. Research progress of flexible piezoresistive strain sensing technology in the field of structural health monitoring[J]. Journal of Building Structures,2024,45(7):242-261

[5]Du Houyi, He Yuxin, Huang Lieran, Gao Ziang, Zhang Ruilin, Liu Hu, Liu Chuntai. Research status of fiber reinforced polymer matrix composites with structural health monitoring function[J]. New Chemical Materials,2025,53(1):9-14

[6]Flah M, Nunez I, Ben Chaabene W, et al. Machine learning algorithms in civil structural health monitoring: A systematic review[J]. Archives of computational methods in engineering, 2021, 28(4): 2621-2643.

[7]Gharehbaghi V R, Noroozinejad Farsangi E, Noori M, et al. A critical review on structural health monitoring: Definitions, methods, and perspectives[J]. Archives of computational methods in engineering, 2022, 29(4): 2209-2235.

[8]Han Q, Ma Q, Xu J, et al. Structural health monitoring research under varying temperature condition: A review[J]. Journal of Civil Structural Health Monitoring, 2021, 11(1): 149-173.

[9]Dong C Z, Catbas F N. A review of

computer vision–based structural health monitoring at local and global levels[J]. Structural Health Monitoring, 2021, 20(2): 692-743.

[10]Bao Y, Li H. Machine learning paradigm for structural health monitoring[J]. Structural health monitoring, 2021, 20(4): 1353-1372.

