International Journal of Advanced Engineering Application (IJAEA)
Volume 3 Issue 7 July 2026
ISSN :3048-6807

Deep Learning and Signal Processing Approaches for Real-Time Damage Detection and Prognosis in Large-Scale Civil Infrastructure: A Comparative Study on a Cable-Stayed Bridge

Lena Hoffmann

Institute for Structural Mechanics, Technical University of Munich, Munich, Germany

Abstract

Structural Health Monitoring (SHM) of large-scale civil infrastructure such as cable-stayed bridges demands continuous acquisition and interpretation of multi-channel sensor data across heterogeneous modalities — accelerometers, fibre-optic strain gauges, acoustic emission transducers, and corrosion probes — generating data volumes that overwhelm traditional signal processing paradigms. This study presents a comparative evaluation of six machine learning architectures — Support Vector Machines (SVM), Random Forest (RF), Long Short-Term Memory networks (LSTM), a convolutional-LSTM hybrid (CNN-LSTM), Autoencoder with multilayer perceptron classifier (AE-MLP), and a Vision Transformer (ViT) adapted for multivariate time-series — applied to a 1.2 km cable-stayed bridge instrumented with 196 sensors over a 36-month monitoring period. Damage scenarios simulated include wire fatigue in hangers, bearing degradation, anchor bolt loosening, and deck delamination across four severity levels (L1–L4). The CNN-LSTM hybrid achieves the highest overall detection accuracy of 97.4% (F1 = 0.974) with a mean time-to-detection of 4.2 minutes for L3 damage events, outperforming the baseline SVM by 6.2 percentage points. Explainability analysis via Gradient-weighted Class Activation Mapping (Grad-CAM) identifies frequency bands 0.3–2.1 Hz and 8.4–12.6 Hz as primary discriminative features for hanger wire fatigue and bearing degradation respectively. A lifecycle-integrated cost model demonstrates that early AI-driven detection of L2 damage reduces maintenance intervention cost by 38% relative to periodic manual inspection schedules, with a net present value improvement of INR 4.2 crore over 30 years.

Keywords: *structural health monitoring, deep learning, LSTM, CNN, cable-stayed bridge, damage detection, vibration-based SHM, explainable AI, Grad-CAM, sensor fusion*

1. Introduction

India's National Infrastructure Pipeline projects over INR 111 lakh crore in infrastructure investment through 2025, of which bridges, viaducts, and elevated corridors constitute a major and safety-critical component. The 2018 Genoa Morandi Bridge collapse and the 2021 Chamoli bridge disaster underscored that conventional visual inspection regimes — typically conducted at three-to-five year intervals by certified engineers — are structurally inadequate to detect subsurface or progressive damage modes before they reach critical severity. Structural Health Monitoring addresses this gap by embedding sensing systems directly into infrastructure, enabling continuous data acquisition and automated condition assessment.

Traditional SHM analysis relies on modal analysis — extracting natural frequencies, mode shapes, and damping ratios from measured acceleration responses and comparing them against baseline finite element models. While effective for global stiffness change detection, modal methods suffer from low sensitivity to local damage at early severity levels and require long data windows for reliable frequency resolution. The emergence of high-capacity edge computing, wireless sensor networks (WSN) with energy harvesting, and deep learning architectures capable of learning discriminative features directly from raw time-series data has created an opportunity to overcome these limitations.

This study addresses three unresolved questions in the literature: (1) which ML architecture offers the best accuracy-latency trade-off for real-time SHM on a multi-modal sensor network with 196 channels; (2) what sensor modalities contribute most to classification performance as revealed by feature importance and Grad-CAM analysis; and (3) what is the lifecycle cost impact of transitioning from periodic manual inspection to AI-driven continuous SHM on a representative cable-stayed bridge in Indian infrastructure conditions.

2. Literature Review

Vibration-based SHM methods have been comprehensively reviewed by Doebling et al. [1] and Farrar & Worden [2], who classify approaches into model-based and data-driven categories. Model-based methods compare measured responses to FE model predictions but are sensitive to modelling uncertainty and require expensive recalibration after geometric or material modifications. Data-driven methods, by contrast, learn damage-sensitive features directly from historical data, making them applicable even when accurate structural models are unavailable.

Convolutional Neural Networks (CNNs) applied to spectrogram representations of acceleration signals have demonstrated strong performance on benchmark datasets such as the Z24 Bridge benchmark [3] and the ASCE structural health monitoring benchmark [4]. However, CNNs process fixed-length windows and do not inherently model temporal dependencies across longer time horizons. LSTM networks, which maintain gated memory cells enabling learning of long-range dependencies, have shown superior performance on non-stationary structural response signals where damage manifests as subtle drift in damping ratios over weeks or months [5,6]. Hybrid CNN-LSTM architectures that extract local spectral features via convolutional layers before feeding temporal context to LSTM cells have emerged as a promising unification [7].

Transformer architectures, originally developed for natural language processing tasks, have recently been adapted for multivariate time-series classification via the Vision Transformer (ViT) paradigm applied to 2D time-frequency representations [8]. Their self-attention mechanism enables explicit modelling of inter-channel dependencies across the full sensor network — a property particularly valuable for cable-stayed bridges where damage in a single hanger manifests as altered vibration patterns in adjacent cables and deck segments. Despite promising results on benchmark datasets, systematic comparison against established LSTM-based methods on real infrastructure deployments remains limited in the peer-reviewed literature.

3. Experimental Setup and Data Acquisition

3.1 Bridge Description and Sensor Network

The study structure is a 1,240-metre asymmetric cable-stayed bridge spanning the Godavari river in Telangana, India, comprising a central pylon of 185 m height, 96 stay cables in harp arrangement, and a composite steel-concrete deck. Constructed in 2011, the bridge carries approximately 42,000 vehicles per day and has experienced progressive hanger wire corrosion in the eastern cable group due to inadequate drainage design — providing a naturally-occurring damage case supplementing the controlled damage scenarios.

The 196-channel sensor network deployed for this study is summarised in Table 1. Sensor data is transmitted via a hybrid WSN — wired for accelerometers and FBG sensors requiring high sampling rates, wireless for temperature, humidity, and corrosion probes — to an on-site edge server (NVIDIA Jetson AGX Orin, 64 GB LPDDR5) where preprocessing and real-time inference are executed.

Table 1. Sensor Network Composition and Deployment Parameters

Sensor Type	Quantity	Sampling Rate (Hz)	Primary Target
Piezoelectric Accelerometers	64	2000	Modal frequencies & damping

Fibre Optic (FBG) Strain	48	100	Strain fields & fatigue
Ultrasonic Transducers	32	5000	Wave propagation / cracks
Acoustic Emission	24	1000 kHz	Crack initiation & growth
Temperature / Humidity	16	1	Environmental correction
Corrosion Potential Probes	12	0.1	Rebar electrochemical state

Table 1 summarises the sensor modalities, quantities, sampling rates, and primary damage targets. The total data throughput is approximately 3.2 GB per hour of continuous monitoring.

3.2 Damage Scenarios and Severity Levels

Four damage scenarios were investigated across four severity levels (L1: incipient, L2: minor, L3: moderate, L4: severe): (i) stay cable wire fatigue — induced by controlled notching of individual wire strands in decommissioned cable segments; (ii) elastomeric bearing degradation — simulated via partial removal of bearing layers; (iii) anchor bolt loosening — torque reduction from 240 Nm to 60, 40, 20, and 10 Nm; and (iv) deck delamination — 0.5, 1.0, 2.0, and 4.0 m² bonded patches detached from the soffit. In addition, the naturally-occurring hanger wire corrosion in the eastern cable group (estimated at 12% cross-sectional area loss by UT inspection) serves as an in-situ validation case.

4. Machine Learning Methods

4.1 Feature Extraction and Preprocessing Pipeline

Raw acceleration signals are preprocessed through a five-stage pipeline: (1) anti-aliasing low-pass filtering at 800 Hz (Butterworth, 8th order); (2) environmental compensation using a multilinear regression model regressing modal frequencies against ambient temperature and traffic loading index, trained on undamaged baseline data; (3) 30-second Hamming-windowed segments with 50% overlap, yielding 192 samples per hour per channel; (4) Welch power spectral density estimation (NFFT = 8192) and continuous wavelet transform (Morlet basis, scales 1–512) for time-frequency representation; and (5) inter-channel coherence matrices computed for the 64-accelerometer subnetwork, providing structural coupling information unavailable from individual channel analysis.

Fig. 1 illustrates the end-to-end data flow from sensor acquisition to real-time damage classification output, with the three principal analytical branches — frequency domain (PSD), time-frequency (CWT scalogram), and cross-channel (coherence matrix) — feeding into the respective ML model input layers.

4.2 Model Architectures

The CNN-LSTM hybrid — the best-performing architecture — processes 2D CWT scalograms through four convolutional blocks (3×3 kernels, batch normalisation, ReLU, max-pooling) before flattening spatial features into a sequence fed to a two-layer bidirectional LSTM (256 hidden units per direction) followed by a dense softmax classifier over the 17 classes (1 undamaged + 4 damage types × 4 severity levels). Total trainable parameters: 4.2 million. Training was performed on 80% of the 36-month dataset (approximately 4.2 million labelled 30-second windows) using AdamW optimiser with cosine learning rate annealing, batch size 64, on a dual NVIDIA A100 GPU cluster, requiring 14 hours for convergence.

The Vision Transformer adapts the ViT-B/16 patch embedding approach to 224×224 CWT scalogram images, dividing each image into 196 non-overlapping 16×16 patches embedded into a 768-dimensional space. Positional encodings and a learnable class token are prepended before 12 multi-head self-attention (MHSA) layers (12 heads, attention dimension 768). Fine-tuning from ImageNet-21k pretrained weights required only 6 hours, with the self-attention mechanism enabling inter-channel dependency capture through concatenation of multi-channel scalograms into a single RGB-equivalent image.

5. Results and Discussion

5.1 Classification Performance

Table 2 presents the overall classification performance of all six algorithms on the 20% held-out test set. The CNN-LSTM hybrid achieves the highest accuracy (97.4%) and F1-score (0.974), outperforming the ViT (96.9%) by a margin that is statistically significant (McNemar's test, $p < 0.01$). The LSTM achieves 95.8% accuracy, confirming that temporal modelling provides substantial uplift over the purely feature-based SVM (91.2%) and Random Forest (93.6%). The Autoencoder-based approach (94.1%) offers a competitive semi-supervised alternative valuable in real deployments where labelled damage data is scarce.

Table 2. Classification Performance Summary Across All Six ML Algorithms

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score	Training Time (s)
SVM (RBF)	91.2	90.4	91.8	0.911	42
Random Forest	93.6	92.9	94.1	0.935	18
LSTM Network	95.8	95.2	96.3	0.957	312
CNN-LSTM Hybrid	97.4	97.1	97.6	0.974	485
Autoencoder + MLP	94.1	93.7	94.4	0.940	228
Transformer (ViT)	96.9	96.5	97.2	0.969	624

Table 2. Test-set performance metrics for all six algorithms trained on 36-month monitoring data. Training time refers to a single epoch on dual A100 GPU. SVM and Random Forest trained on CPU (Intel Xeon 6226R).

Fig. 2 visualises the confusion matrices and severity-stratified detection rates. Panel A shows that the CNN-LSTM hybrid achieves near-perfect classification for L3 and L4 damage events (>99.1% across all damage types) while showing the highest misclassification rate between L1 and undamaged states — expected given that incipient damage induces sub-threshold modal parameter shifts. Panel B illustrates how detection rate varies with severity level across all algorithms, confirming that L1 detection remains the primary differentiator between methods: CNN-LSTM achieves 89.3% L1 detection versus 71.6% for SVM. Panel C presents ROC curves for the binary undamaged/damaged classification task, with CNN-LSTM achieving AUC = 0.998.

5.2 Explainability and Feature Attribution

Grad-CAM activation maps computed for correctly-classified L3 hanger fatigue events reveal consistent high activation in the 0.3–2.1 Hz frequency band corresponding to the first four bending modes of the cable array, and in the 8.4–12.6 Hz range associated with local wire resonance frequencies. Bearing degradation events show dominant activation at 15–22 Hz, consistent with increased friction-induced high-frequency excitation. These results confirm physical interpretability of the CNN-LSTM's learned representations rather than reliance on spurious correlations, addressing a key concern in safety-critical infrastructure AI deployment.

Fig. 3 presents the Grad-CAM visualisations alongside the Random Forest feature importance profile for comparison. Panel A overlays Grad-CAM activation regions on representative CWT scalograms for all four damage types at L3 severity. Panel B shows that accelerometers positioned at the quarter-span deck locations (nodes A12, A28, A44, A60) and mid-cable transducers (nodes C08, C24) account for 67.3% of total Random Forest feature importance, confirming that sensor placement optimisation could reduce the network from 64 to approximately 22 accelerometers while retaining 94%+ accuracy — with significant implications for deployment cost.

5.3 Lifecycle Cost Analysis

A 30-year lifecycle cost model comparing periodic manual inspection (current practice: bi-annual, INR 18 lakh per campaign) against AI-driven continuous SHM was developed using a Monte Carlo simulation framework (10,000 iterations) accounting for uncertainty in damage growth rates and intervention costs. The model demonstrates that AI-driven detection of L2 damage reduces expected maintenance intervention cost by 38% through avoidance of reactive repair campaigns triggered by L4 damage events, which carry a 4.7× cost premium over preventive intervention at L2. The net present value of transitioning to continuous SHM over 30 years, accounting for sensor installation (INR 2.8 crore), annual maintenance (INR 32 lakh/year), and edge computing infrastructure (INR 85 lakh), is estimated at INR 4.2 crore positive NPV at a 7% discount rate representative of Indian infrastructure financing conditions.

6. Conclusion

This study demonstrates that hybrid CNN-LSTM architectures applied to continuous wavelet transform scalogram representations of multi-channel structural response data achieve state-of-the-art damage detection performance (97.4% accuracy, F1 = 0.974) on a 196-sensor cable-stayed bridge monitoring network over a 36-month deployment period. Key conclusions are:

1. CNN-LSTM outperforms all compared architectures, including Vision Transformer (ViT), LSTM, SVM, Random Forest, and AE-MLP, with statistically significant margins at both overall accuracy and incipient (L1) damage detection rate (89.3% vs 71.6% for SVM).
2. Grad-CAM explainability analysis identifies physically meaningful discriminative frequency bands (0.3–2.1 Hz for hanger fatigue; 8.4–12.6 Hz for cable resonance; 15–22 Hz for bearing degradation), establishing interpretability sufficient for safety-critical deployment.
3. Sensor importance analysis indicates that an optimised 22-accelerometer subnetwork retains 94%+ accuracy, suggesting that future deployments can reduce sensor count by 66% without commensurate performance degradation.
4. The lifecycle cost model yields a positive NPV of INR 4.2 crore over 30 years relative to conventional bi-annual inspection, driven primarily by early detection enabling preventive rather than reactive maintenance interventions.

Future work will address federated learning approaches enabling model updating from multiple bridge deployments without centralised data sharing — critical for privacy-preserving national infrastructure monitoring networks — and physics-informed neural network (PINN) integration to enforce structural mechanics constraints in damage state estimation.

References

[1] Doebling, S. W., Farrar, C. R., Prime, M. B., & Shevitz, D. W. (1996). Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: a literature review. Los Alamos National Laboratory Report LA-13070-MS.

[2] Farrar, C. R., & Worden, K. (2012). Structural Health Monitoring: A Machine Learning Perspective. John Wiley & Sons.

[3] Maeck, J., & De Roeck, G. (1999). Dynamic bending and torsion stiffness derivation from modal curvatures and torsion rates. Journal of Sound and Vibration, 225(1), 153–170.

[4] Johnson, E. A., Lam, H. F., Katafygiotis, L. S., & Beck, J. L. (2004). Phase I IASC-ASCE structural health monitoring benchmark problem using simulated data. Journal of Engineering Mechanics, 130(1), 3–15.

[5] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.

[6] Avci, O., Abdeljaber, O., Kiranyaz, S., Hussein, M., Gabbouj, M., & Inman, D. J. (2021). A review of vibration-based damage detection in civil structures. Mechanical Systems and Signal Processing, 147, 107077.

[7] Zhang, W., Peng, G., Li, C., Chen, Y., & Zhang, Z. (2017). A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. Sensors, 17(2), 425.

[8] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. ICLR 2021.

[9] Pathirage, C. S. N., Li, J., Li, L., Hao, H., Liu, W., & Ni, P. (2018). Structural damage identification based on autoencoder neural networks and deep learning. Engineering Structures, 172, 13–28.

[10] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. ICCV 2017, 618–626.

[11] Sun, L., Shang, Z., Xia, Y., Bhowmick, S., & Nagarajaiah, S. (2020). Review of bridge structural health monitoring aided by big data and artificial intelligence. Journal of Structural Engineering, 146(11), 04020280.

[12] Sohn, H., Farrar, C. R., Hemez, F. M., Shunk, D. D., Stinemates, D. W., Nadler, B. R., & Czarnecki, J. J. (2003). A review of structural health monitoring literature: 1996–2001. Los Alamos National Laboratory Report LA-13976-MS.