Proc. of the European Safety and Reliability Conference (ESREL 2026)
Edited by José C. Matos, Paulo B. Lourenço, Michael Beer, Daniel Oliveira, Edoardo Patelli, Hélder S. Sousa, Rui A. Silva and Monica Santamaria

doi: 10.3850/ESREL2026061419_esrel26-p28954-cd

Semi-Supervised LSTM–Deep Autoencoder Framework for Anomaly Detection and Resilient Predictive Maintenance in Bridge Infrastructure

Jabez Nesackon Abraham

ISISE, ARISE, Department of Civil Engineering, University of Minho, Portugal. E-mail:id10779@alunos.uminho.pt

Jose C. Matos

ISISE, ARISE, Department of Civil Engineering, University of Minho, Portugal. E-mail: jmatos@civil.uminho.pt

Son Dang Ngoc

ISISE, ARISE, Department of Civil Engineering, University of Minho, Portugal. E-mail: sondn@civil.uminho.pt

Jerusha Samuel Jayaraj

Department of Information Systems, University of Minho, Portugal. E-mail: id12207@alunos.uminho.pt

Maria Rosa Valluzzi

Department of Cultural Heritage, University of Padova, Italy. E-mail: mariarosa.valluzzi@unipd.it

Detecting early signs of structural damage is vital for the safety of engineering assets, but natural environmental and operational changes often hide these subtle warning signs. To address this challenge, this study introduces Deep Autoencoder (DAE) and Attention-based Long Short-Term Memory Autoencoder (ALSTM-AE), a semi-supervised machine learning framework designed for Structural Health Monitoring (SHM). The proposed method operates by processing raw vibration signals from a structure and extracting key features that capture its natural movement. It then uses a two-part neural network: a DAE learns the fundamental patterns of the structure's normal behavior, while an ALSTM-AE tracks how those patterns change over time. The system flags potential damage by calculating an anomaly score, which measures how much the current data deviates from normal expectations, and applying a strict mathematical threshold, median absolute deviation (MAD), to make the final decision. When tested on the Z24 bridge dataset, the proposed framework significantly outperformed traditional methods across key metrics, including accuracy, precision, and recall. The results demonstrate that this new approach is highly effective at ignoring normal, load-induced variations while accurately catching real damage, providing a strong foundation for real-time monitoring, predictive maintenance, and digital twin applications.

Keywords: SHM; Vibration-based monitoring; Anomaly Detection; Semi-supervised Learning; Principal Component Analysis; Autoencoder; LSTM-DeepAE; Predictive Maintenance.

1. Introduction

Structural health monitoring (SHM) is a vital technology for maintaining the safety, reliability, and longevity of critical infrastructure systems, including bridges, buildings, and transportation networks (Farrar and Worden 2012; Worden et al. 2007). Identifying subtle structural anomalies from monitoring data remains challenging due to environmental and operational variability, measurement noise, and the limited availability of labeled damaged-state data (Farrar and Worden 2006; Worden and Manson 2006). Recent studies further emphasize that the scarcity of labeled structural damage data remains a major obstacle for the development of reliable data-driven SHM systems (Bao et al. 2025; Jia and Li 2023).

To overcome these limitations, machine learning techniques have been increasingly utilized in SHM applications (Worden and Manson 2007). Machine learning methods can automatically learn patterns from large-scale monitoring data and identify abnormal structural behavior. Classical statistical approaches such as

Principal Component Analysis (PCA), clustering methods, and statistical control charts have been widely used for anomaly detection in structural monitoring (Hastie et al. 2004). PCA-based approaches are inherently linear and may struggle to model the complex nonlinear relationships present in structural vibration data.

Recent advances in artificial intelligence and deep learning have introduced powerful data-driven techniques capable of learning nonlinear representations from high-dimensional sensor data (Cha et al. 2024; Ye et al. 2019). Autoencoders have emerged as an effective unsupervised learning framework for anomaly detection in SHM systems by learning compressed latent representations of normal structural behavior (Hinton and Salakhutdinov 2006). Recurrent neural networks, particularly Long Short-Term Memory (LSTM) networks, have demonstrated strong capabilities in modeling sequential patterns in time-series data, enabling improved detection of subtle temporal variations in structural responses (Malhotra et al. 2016; Semper et al. 2025). Recent studies have investigated hybrid frameworks that integrate statistical feature extraction with deep learning models to enhance robustness and reliability in anomaly detection tasks (Abraham et al. 2026; Benfenati et al. 2024; Singh et al. 2026).

To address these challenges, this paper proposes a hybrid semi-supervised anomaly detection framework that integrates classical statistical modeling with deep representation learning. A PCA-based autoencoder framework is used as a baseline model, and to capture more complex nonlinear patterns and temporal dependencies, a deep autoencoder combined with an attention-enhanced LSTM is developed to learn latent representations of normal structural behavior from baseline data. The effectiveness of the proposed framework is evaluated using the widely studied Z24 Bridge monitoring dataset (Reynders and Roeck 2008), which provides long-term vibration measurements collected during controlled damage scenarios.

2. Proposed Methodology

In real-world SHM, obtaining labeled data for damaged structures is a major obstacle, whereas baseline data from healthy structures is abundant. To address this data scarcity, the proposed methodology employs a semi-supervised learning strategy. The proposed pipeline leverages feature-based PCA, a Deep Autoencoder (DAE), and an Attention-based LSTM Autoencoder (ALSTM-AE) applied to the latent representations of the

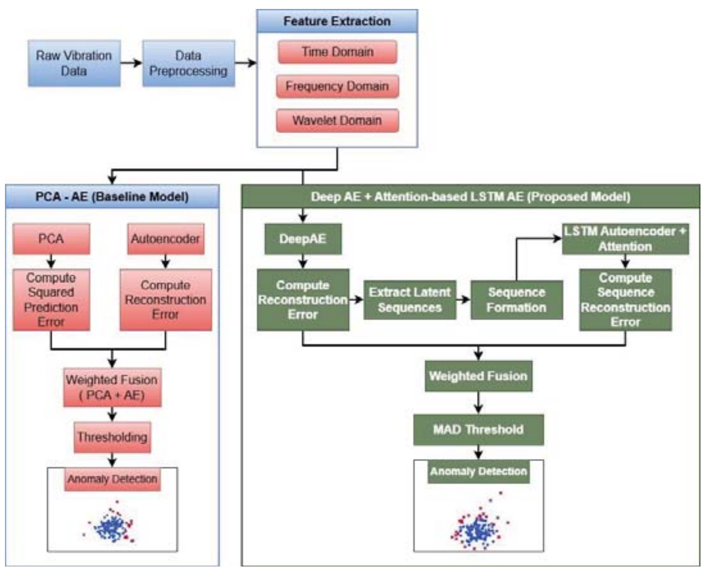


Fig. 1. Workflow of the semi-supervised anomaly detection method

DAE. The final anomaly score is computed as a weighted fusion of DAE reconstruction error and ALSTM-AE latent sequence error, while PCA is used as a baseline comparator. The overall workflow is shown in Figure 1.

2.1. *Baseline Model: PCA + Autoencoder*

In previous studies (Abraham et al. 2026), the anomaly detection framework relied on a hybrid PCA and shallow AE applied to physics-informed features derived from multichannel vibration signals. PCA captured the dominant variance in the feature space, and deviations from the learned subspace were used as anomaly indicators. The shallow AE was trained on baseline (healthy state) features to model nonlinear interactions, with reconstruction errors serving as an additional anomaly score.

Limitations of the baseline PCA + AE approach:

- **Limited Nonlinear Representation**: Shallow autoencoders capture only low-complexity nonlinear interactions, restricting their ability to model intricate structural behaviors in large-scale bridges.
- **No Temporal Modeling**: PCA + AE operates on individual feature vectors for each window, ignoring sequential correlations in vibration signals that are critical for detecting slowly evolving or subtle anomalies.
- **Sensitivity to Environmental Variability**: The baseline model may produce false positives under temperature or load-induced modal variations because it lacks mechanisms to account for temporal patterns or environmental context.
- **Fixed Feature Representation**: PCA relies on linear projections of the features, which may overlook higher-order dependencies between spectral, statistical, and wavelet-based features.

These limitations highlight the need for a more expressive, sequence-aware anomaly detection framework capable of capturing nonlinear temporal dependencies in structural vibration data.

2.2. *Proposed Method: DAE + ALSTM-AE*

To address the limitations of the PCA + AE baseline, the proposed framework introduces two key enhancements by integrating deeper nonlinear representation learning and temporal modeling. A DAE extends the shallow architecture into a multi-layer network, enabling richer feature extraction, thereby improving sensitivity to subtle anomalies. In addition, an ALSTM-AE model captures temporal dependencies by processing sequences of latent vectors from the DAE, using multi-head attention to focus on informative time steps and better detect evolving structural changes while reducing false positives caused by environmental variability.

To identify anomalies, the system calculates a fused anomaly score (s_{fused}) by combining the reconstruction errors of DAE (s_{DAE}) and ALSTM-AE ($s_{ALSTM\text{-}AE}$). This final score uses empirically derived weighting coefficients to balance the outputs:

$$s_{fused} = \alpha \cdot s_{DAE} + \beta \cdot s_{ALSTM-AE} \qquad (1)$$

Based on experimental optimization, the spatial DAE error is weighted by α=0.4, while the temporal ALSTM-AE sequence error is weighted by β=0.6. Finally, a robust statistical thresholding strategy utilizing the median and median absolute deviation (MAD) is applied to this fused score to separate normal variations from true anomalies and make the final decision.

Overall, this hybrid framework combines spatial and temporal anomaly modeling, improving detection performance while maintaining robustness, as demonstrated on the Z24 bridge dataset.

3. Case Study

3.1. *Z24 Bridge Dataset*

The experimental evaluation was conducted using the publicly available Z24 Bridge dataset, which is widely used in structural health monitoring and anomaly detection research. The Z24 Bridge, a prestressed concrete highway structure located in Switzerland, underwent extensive monitoring prior to its demolition as part of a comprehensive long-term structural health monitoring campaign. The dataset consists of multichannel vibration signals collected from accelerometers placed at various locations on the bridge deck. These sensors continuously recorded structural responses under various environmental and operational conditions, including traffic loads, temperature variations, and artificial damage scenarios.

3.2. *Quantitative Results and Performance Analysis*

To assess the effectiveness of the proposed anomaly detection framework, various performance metrics and visual analyses were performed. The results compare the baseline model (PCA + AE) with the proposed model (DAE + ALSTM-AE) on structural vibration data from the Z24 Bridge dataset. The evaluation includes quantitative metrics and graphical analyses, such as Receiver Operating Characteristic (ROC) curves, Precision–Recall (PR) curves, reconstruction error distributions, training loss curves, and comparative performance charts.

3.2.1. *Quantitative Performance Comparison*

Table 1 provides a comprehensive comparison of the overall performance between the baseline model and the proposed model, utilizing standard classification metrics.

Table 1. Performance Comparison.

Metric	PCA + AE	DAE + ALSTM-AE
AUC	0.923	0.945
F1 Score	0.892	0.912
Precision	0.964	0.973
Recall	0.830	0.859
Accuracy	0.832	0.863

The results clearly demonstrate that the proposed architecture consistently outperforms the baseline model across all evaluation metrics. The most notable improvements are observed in Area Under the Curve (AUC), recall, and overall accuracy, indicating enhanced capability in detecting anomalous structural behaviors. The improvement in recall is particularly important for SHM applications, as it reflects the ability of the model to correctly identify abnormal structural conditions.

3.2.2. *ROC Curve Analysis*

The ROC curve presented in Figure 2 illustrates the relationship between the True Positive Rate (TPR) and the False Positive Rate (FPR) across various thresholds for anomaly scores. The ROC curves derived from the experiments demonstrate that the proposed model consistently achieves a higher true positive rate at any given false positive rate compared to the baseline method. The AUC for the proposed model is 0.945, surpassing the AUC of 0.923 for the baseline model.

A higher AUC value indicates stronger discriminative capability between normal and anomalous structural states. The improvement observed in the proposed framework demonstrates that incorporating deep representation learning and temporal modeling significantly enhances anomaly detection performance.

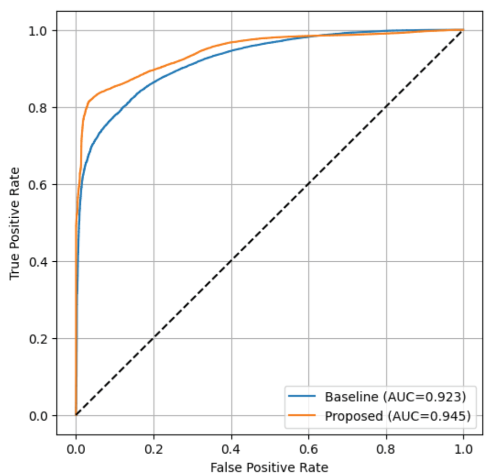


Fig. 2. ROC Curve comparison between baseline and proposed models.

3.2.3. *PR Curve Analysis*

In anomaly detection problems, datasets are often highly imbalanced because anomalous events occur less frequently than normal conditions. Under such circumstances, the PR curve provides a more informative evaluation than ROC analysis. Figure 3 shows the PR curve generated from the experiments, which shows that the proposed model maintains higher precision across a wide range of recall values compared to the baseline approach. This indicates that the proposed model can detect anomalies more accurately while maintaining a lower false alarm rate.

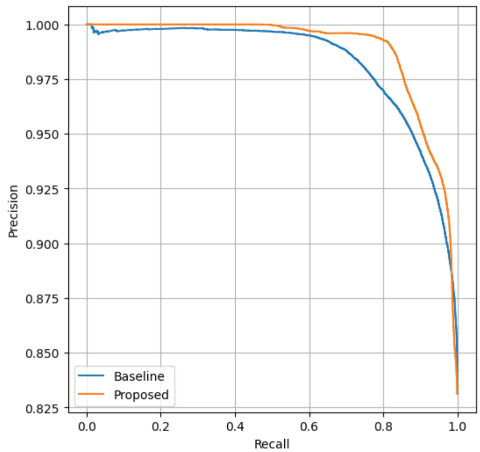


Fig. 3.PR curve comparison between baseline and proposed models.

3.2.4.*Reconstruction Error Distribution*

Reconstruction-based anomaly detection methods are founded on the principle that models trained using normal data will demonstrate low reconstruction errors for typical samples, while producing higher reconstruction errors for samples with anomalies. Figure 4 shows the distribution of reconstruction errors for both normal and anomalous data samples. The distribution plot shows that anomalous samples generally produce significantly higher reconstruction errors than normal samples. A threshold derived from robust statistical estimation (median and MAD) separates the two distributions. This clear separation between normal and anomaly reconstruction errors confirms the effectiveness of the proposed reconstruction-based anomaly scoring mechanism.

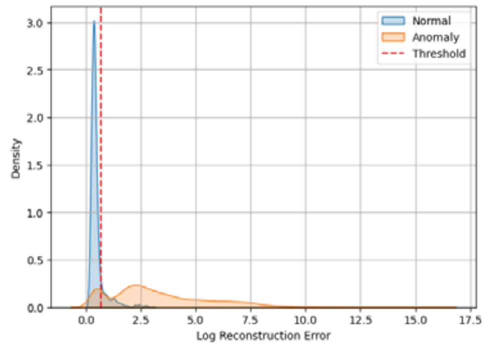
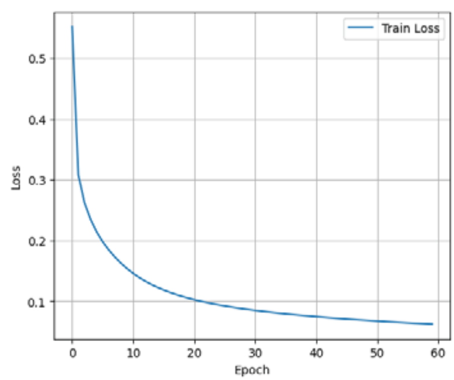


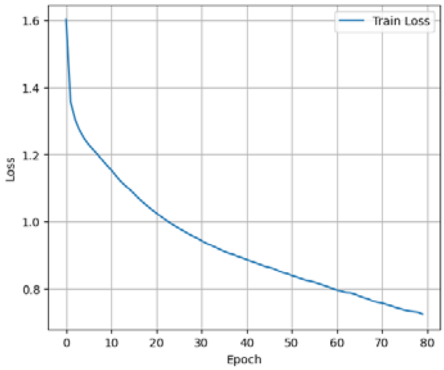
Fig. 4.Reconstruction error distribution.

3.2.5.*Training Loss Curves*

Training loss curves were analyzed to verify the convergence behavior of the deep learning models. The DAE loss curve in Figure 5(a) shows a rapid decrease during the initial training epochs, followed by gradual stabilization as the training progresses. This indicates that the model successfully learns meaningful representations of structural vibration patterns. Similarly, the ALSTM-AE loss curve in Figure 5(b) shows stable convergence, indicating that the temporal sequence model effectively captures dependencies across sequential vibration windows. The smooth decline in the loss curves, with no significant oscillations, indicates that the training process is stable and that the models are not suffering from severe overfitting.



(a)



(b)

Fig. 5.Training loss curve: (a) DAE, (b) ALSTM-AE.

4. Discussion

The experimental results demonstrate that the proposed DAE + ALSTM-AE framework significantly outperforms the baseline PCA–Autoencoder approach for anomaly detection in structural vibration signals.

The results highlight several key factors contributing to the effectiveness of the proposed framework for structural anomaly detection. Traditional approaches, such as PCA, rely on linear dimensionality reduction and often fail to capture complex nonlinear relationships in structural vibration data. In contrast, the deep autoencoder employed in this study learns compact nonlinear feature representations through multiple hidden layers, enabling the extraction of subtle patterns and correlations that improve anomaly detection performance, particularly in terms of AUC and F1-score.

Structural vibration signals exhibit inherent temporal dependencies that are not captured by baseline models processing individual windows independently. The integration of an LSTM Autoencoder addresses this limitation by modeling sequential patterns and learning temporal dynamics of normal structural behavior, allowing deviations to be identified through increased reconstruction errors.

The incorporation of an attention mechanism further enhances this capability by assigning greater importance to informative time steps, enabling the model to focus on critical segments of the signal and better distinguish between normal variations and true anomalies. Moreover, analysis of reconstruction errors shows a clear separation between normal and anomalous samples, validating their use as a reliable anomaly indicator. The adoption of robust thresholding using the median and median absolute deviation (MAD) further improves stability by reducing the influence of noise and outliers. Overall, the proposed framework demonstrates strong potential for real-world SHM applications by enabling continuous, automated detection of structural anomalies without requiring labeled damage data, thereby supporting early warning systems and facilitating timely maintenance to enhance infrastructure safety.

While the results are promising, there are several limitations to consider. The framework has only been evaluated on a single benchmark dataset, and it is essential to conduct further testing on additional datasets to thoroughly assess its generalization capabilities. The reliance on reconstruction-based anomaly detection may also lead to false positives under significant environmental variations. Additionally, the calculation expenses associated with deep learning models may constrain their implementation in real-time monitoring systems that operate with limited resources. Addressing these challenges will be a significant area of focus in future research endeavors.

5. Conclusion

This study presented an effective anomaly detection framework for Structural Health Monitoring (SHM) based on Deep Autoencoders (DAE), LSTM Autoencoders, and attention mechanisms (LSTM-Attention). The proposed method aims to improve anomaly detection performance in multichannel vibration signals collected from the Z24 bridge benchmark dataset.

The framework first extracts multi-domain features from vibration signals and learns compact nonlinear representations using a DAE. These latent representations are then processed by an LSTM Autoencoder with attention (ALSTM-AE) to capture temporal dependencies across sequential vibration windows. Anomaly scores are obtained using reconstruction errors and combined using a robust fusion strategy.

Experimental evaluation on the Z24 Bridge dataset demonstrates that the proposed method consistently outperforms the baseline PCA–Autoencoder model. The proposed approach achieves higher performance across multiple evaluation metrics, including AUC, F1-score, recall, precision, and overall accuracy. The results confirm that integrating deep representation learning, temporal modeling, and attention mechanisms significantly enhances anomaly detection capabilities for structural vibration analysis.

Future research can extend the framework to multi-modal structural monitoring data, including temperature, strain, and acoustic emission signals, to provide a more comprehensive understanding of structural health conditions.

Acknowledgement

This work is financed by national funds through FCT – Foundation for Science and Technology, under grant agreement PRT/BD/154386/2023 attributed to the 1st author. Jabez Nesackon Abraham acknowledges the support of the doctoral grant reference PRT/BD/154386/2023, financed by the Portuguese Foundation for Science and Technology (FCT). under the MIT Portugal Program (2022 MPP2030-FCT). This work was also supported by FCT/MCTES under the R&D Unit Institute for Sustainability and Innovation in Structural Engineering (ISISE), under reference UID/4029/2025, and under the Associate Laboratory Advanced Production and Intelligent Systems ARISE under reference LA/P/0112/2020.

References

Abraham, Jabez Nesackon, Minh Q. Tran, Jerusha Samuel Jayaraj, Jose C. Matos, Maria Rosa Valluzzi, and Son N. Dang. (2026). "Unsupervised Learning-Based Anomaly Detection for Bridge Structural Health Monitoring: Identifying Deviations from Normal Structural Behaviour." Sensors 26 (2).

Bao, Yuequan, Huabin Sun, Yang Xu, Xiaoshu Guan, Qiuyue Pan, and Dawei Liu. (2025). "Recent Advances in Structural Health Diagnosis: A Machine Learning Perspective." Advances in Bridge Engineering 6 (1): 7.

Benfenati, Luca, Daniele Jahier Pagliari, Luca Zanatta, et al. (2025). "Foundation Models for Structural Health Monitoring." IEEE Transactions on Sustainable Computing 10 (6): 1103–15.

Cha, Young-Jin, Rahmat Ali, John Lewis, and Oral Büyüköztürk. (2024). "Deep Learning-Based Structural Health Monitoring." Automation in Construction 161 (May): 105328.

Farrar, Charles R., and Nick A. J. Lieven. (2006). "Damage Prognosis: The Future of Structural Health Monitoring." Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 365 (1851): 623–32.

Farrar, Charles R., and Keith Worden. (2012). Structural Health Monitoring: A Machine Learning Perspective. 1st ed. Wiley.

Hastie, Trevor, Robert Tibshirani, Jerome Friedman, and James Franklin. (2004). "The Elements of Statistical Learning: Data Mining, Inference, and Prediction." Math. Intell. 27 (November): 83–85.

Hinton, G. E., and R. R. Salakhutdinov. (2006). "Reducing the Dimensionality of Data with Neural Networks." Science 313 (5786): 504–7.

Jia, Jing, and Ying Li. (2023). "Deep Learning for Structural Health Monitoring: Data, Algorithms, Applications, Challenges, and Trends." Sensors 23 (21).

Malhotra, Pankaj, Anusha Ramakrishnan, Gaurangi Anand, Lovekesh Vig, Puneet Agarwal, and Gautam Shroff. (2016). LSTM-Based Encoder-Decoder for Multi-Sensor Anomaly Detection. July 1.

Reynders, Edwin, and Guido De Roeck. (2009). "Continuous Vibration Monitoring and Progressive Damage Testing on the Z24 Bridge." In Encyclopedia of Structural Health Monitoring. John Wiley & Sons, Ltd.

Semper, Marc, Manuel Curado, Jose Luis Oliver, and Jose F. Vicent. (2025). "Noise Pollution Prediction in a Densely Populated City Using a Spatio-Temporal Deep Learning Approach." Applied Sciences 15 (10).

Singh, Jaskaran, Meenu Gupta, and Rakesh Kumar, eds. (2026). Smart Technologies and Intelligent Computing. CRC Press.

Worden, Keith, Charles R. Farrar, Graeme Manson, and Gyuhae Park. (2007). "The Fundamental Axioms of Structural Health Monitoring." Proceedings of the Royal Society A:

Mathematical, Physical and Engineering Sciences 463 (2082): 1639–64.

Worden, Keith, and Graeme Manson. (2006). "The Application of Machine Learning to Structural Health Monitoring." Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 365 (1851): 515–37.

Ye, X. W., Jin Tao, and Chung Bang Yun. (2019). "A Review on Deep Learning-Based Structural Health Monitoring of Civil Infrastructures." Smart Structures and Systems 24 (November): 567–85.