Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

Data-Driven Damage Detection in Multi-Storey RC Frame Buildings Using Machine Learning and Multi-Domain Vibration Features

Rohan Saini
Department of Civil Engineering, PKG College of Engineering & Technology, Panipat, Kurukshetra University, India
Corresponding author: Rohan Saini

Abstract - Structural health monitoring (SHM) of civil infrastructure has traditionally relied on visual inspection, which is subjective, labour-intensive and unsuited to continuous, large-scale assessment. This study presents a data-driven vibration-based damage detection framework for a benchmark G+5 reinforced-concrete (RC) moment-resisting frame building. A validated three-dimensional finite element model is used to simulate a healthy baseline condition and seven single- and multi-element damage scenarios (10-40% stiffness reduction). Modal and time-history dynamic analyses generate synthetic acceleration response data at four sensor locations, from which a 42-feature multi-domain set spanning time-domain statistics, frequency-domain modal parameters and Discrete Wavelet Transform (DWT) energy coefficients is extracted. Four classifiers - Support Vector Machine (SVM), Random Forest (RF), Artificial Neural Network (ANN) and k-Nearest Neighbour (k-NN) - are trained and compared under 5-fold cross-validation and under 5% and 10% Gaussian measurement noise. Random Forest achieved the highest performance, with 96.4% test accuracy on clean data and 91.8% accuracy under 10% noise, outperforming ANN (94.8%/89.5%), SVM (91.7%/85.2%) and k-NN (88.3%/79.8%). The combined multi-domain feature set outperformed any single feature category by at least 7.0 percentage points, and natural frequency, mode shape curvature and mid-level wavelet energy emerged as the most discriminative features. The results establish an accurate, noise-resilient and computationally efficient framework for automated SHM of multi-storey RC buildings, with particular relevance to seismically active urban regions.

Keywords: Structural Health Monitoring; Machine Learning; Vibration-Based Features; Random Forest; Damage Detection

1. INTRODUCTION

Civil infrastructure is subjected throughout its service life to environmental exposure, material ageing, fatigue and extreme events such as earthquakes, all of which can produce progressive, often hidden, structural damage. Conventional condition assessment relies on periodic visual inspection, which is costly, subjective and incapable of detecting internal damage or providing continuous coverage. These limitations have motivated extensive research into Structural Health Monitoring (SHM), particularly vibration-based approaches grounded in the principle that localized stiffness loss alters a structure's natural frequencies, mode shapes and other dynamic characteristics.

The growth of machine learning (ML) has enabled data-driven SHM systems capable of learning complex, non-linear relationships between vibration features and damage states without requiring an explicit physics-based inverse model. This is particularly relevant in India, where rapid urbanization has produced a large inventory of multi-storey RC buildings, many located in seismically active zones and constructed prior to modern ductile-detailing provisions. A cost-effective, automated damage-detection framework therefore carries direct practical value for post-earthquake triage and condition-based maintenance.

Despite substantial progress, persistent challenges limit the field deployment of vibration-based ML-SHM: subtle early-stage damage produces frequency shifts of only 1-2%; labelled data from physically damaged real structures is impractical to obtain; there is no clear consensus on the optimal feature-classifier combination for multi-storey RC frames; and few studies systematically evaluate classifier robustness under controlled, varying measurement noise. This study addresses these gaps through a controlled, multi-domain feature and multi-classifier comparison, evaluated under both clean and noise-contaminated conditions on a representative G+5 RC benchmark structure.

The specific objectives are to: (i) develop a validated 3D finite element model of a G+5 RC moment-resisting frame; (ii) simulate healthy, single-element and multi-element damage scenarios at multiple severities; (iii) generate synthetic vibration data through modal and dynamic analysis; (iv) extract time-, frequency- and wavelet-domain features; (v) train and rigorously compare SVM, Random Forest, ANN and k-NN classifiers; (vi) evaluate robustness under simulated measurement noise; and (vii) identify the most discriminative features and provide practical recommendations for sensor deployment and classifier selection.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

2. LITERATURE REVIEW

2.1 Vibration-Based Damage Detection

Vibration-based SHM rests on the principle that stiffness reduction alters measurable dynamic characteristics. Early work by Cawley and Adams established frequency-ratio methods for damage location, while Doebling et al. formalized the widely used four-level SHM hierarchy of detection, localization, quantification and prognosis. Pandey et al. showed that mode shape curvature offers substantially better spatial localization than raw frequencies or mode shapes, a finding incorporated directly into the present feature set. Subsequent reviews (Fan and Qiao; Carden and Fanning) confirmed a shift toward data-driven methods that avoid the need for a calibrated FE baseline, and highlighted environmental variability and automated interpretation as persistent open challenges.

2.2 Machine Learning in SHM

Machine learning has been applied to SHM since the late 1990s, beginning with ANN-based novelty detection (Worden et al.) and statistical pattern recognition combined with autoregressive features (Sohn et al.). SVM classifiers were shown to perform well with limited training data (Yan et al.), while later reviews (Rafiei and Adeli; Avci et al.) concluded that well-engineered, feature-based shallow classifiers - particularly Random Forest - remain highly competitive with, and in many cases superior to, deep learning approaches when datasets are moderate in size. Building-specific case studies (Lin and Zhu; Khoshnoudian et al.; Sun et al.) confirm the feasibility of ML-based vibration SHM for multi-storey structures while highlighting sensitivity to sensor configuration and damage severity.

2.3 Feature Extraction Approaches

Time-domain statistical descriptors (RMS, kurtosis, skewness) capture impulsive and non-stationary signal characteristics (Nair et al.). Frequency-domain features, particularly natural frequencies and modal parameters, remain the most physically interpretable but offer limited spatial resolution and sensitivity to small damage (Salawu). Wavelet-based features, using multi-resolution decomposition, have been shown to localize damage-induced non-stationarities not readily visible in pure time- or frequency-domain representations (Staszewski; Sun and Chang), motivating their inclusion in the present multi-domain pipeline.

2.4 Research Gap

The reviewed literature reveals four specific gaps: (i) most studies use simplified beam, truss or laboratory-scale structures rather than realistic multi-storey RC frames; (ii) few studies systematically compare individual and combined time-, frequency- and wavelet-domain feature categories under a controlled protocol; (iii) noise robustness is rarely evaluated quantitatively; and (iv) comparative classifier benchmarks frequently use inconsistent feature sets or evaluation protocols. This study addresses all four gaps using a common G+5 RC benchmark, a common 42-feature multi-domain set, an explicit noise-sensitivity analysis, and identical training/evaluation protocols across four classifiers.

3. RESEARCH METHODOLOGY

3.1 Benchmark Structure and Finite Element Model

The benchmark structure is a regular, symmetric G+5 RC moment-resisting frame with six floor levels, 3.0 m uniform storey height (18.0 m total height), 15 m x 15 m plan dimensions and three bays at 5 m spacing in each direction. Columns are 400 x 400 mm, beams 250 x 450 mm, slab thickness 125 mm, with M25 concrete and Fe415 reinforcement. The structure is modelled as a fixed-base 3D space frame with rigid floor diaphragms. The fundamental period from FE modal analysis was validated against the IS 1893 (Part 1):2016 empirical estimate, with agreement within 5%, and modal mass participation exceeding 90% in both principal directions.

3.2 Damage Scenarios

Damage is idealized as a localized flexural stiffness reduction. Eight conditions are simulated: a healthy baseline (D0); single-element beam damage at 10%, 20% and 30% severity (D1-D3); single-element column damage at 20% and 40% severity (D4-D5); and two multi-element scenarios combining simultaneous beam and column damage (D6-D7). This design spans varying severity at a fixed location, varying location at comparable severity, and distributed multi-element damage.

3.3 Synthetic Vibration Data Generation

Ambient excitation is modelled as band-limited Gaussian white noise (0-30 Hz) applied at the base in both horizontal directions, consistent with operational modal analysis practice. Sixty-second acceleration records at 200 Hz are generated at four sensor locations (column tops at storeys 2, 3, 4 and 6). For each of the eight damage scenarios, 200 samples are generated through controlled randomization of excitation characteristics, yielding a balanced dataset of 1600 samples.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

3.4 Multi-Domain Feature Extraction

Seven time-domain statistics (RMS, standard deviation, kurtosis, skewness, crest factor, peak-to-peak, mean absolute value) are computed per sensor across four sensors (28 features). Nine frequency-domain features comprise the first five natural frequencies and four mode-shape-curvature values at sensor locations. Five wavelet-domain features are obtained from Daubechies-4 (db4), five-level DWT energy decomposition. The combined 42-feature vector is z-score normalized using training-set statistics prior to classification.

3.5 Classifiers and Evaluation Protocol

Four classifiers are trained and tuned via 5-fold stratified cross-validation on an 80/20 stratified train-test split (1280/320 samples): SVM with RBF kernel (C=10, gamma=0.01); Random Forest (200 trees, unconstrained depth); a feed-forward ANN (42-64-32-8 architecture, ReLU/softmax, Adam optimizer, dropout 0.3, early stopping); and k-NN (k=7, Euclidean distance). Performance is quantified using macro-averaged accuracy, precision, recall and F1-score. Robustness is assessed by re-extracting features after injecting 5% and 10% additive Gaussian measurement noise into the acceleration signals.

4. RESULTS AND DISCUSSION

4.1 Natural Frequency Trends

Natural frequencies decreased monotonically with increasing damage severity, confirming the physical validity of the damage implementation. The fundamental frequency reduced from 2.841 Hz (healthy) to 2.796 Hz for the lowest-severity case (D1, 10% beam damage), a shift of only about 1.6%, illustrating the inherent difficulty of detecting early-stage damage from frequency alone. Column damage produced proportionally larger frequency reductions than comparable-severity beam damage, consistent with the greater contribution of columns to overall lateral stiffness.

4.2 Classifier Performance

Table 1 summarizes macro-averaged performance on the held-out test set (320 samples).

Table 1. Classifier performance comparison (clean data)

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	96.4	96.1	96.4	96.2
ANN (MLP)	94.8	94.5	94.8	94.6
SVM (RBF)	91.7	91.3	91.7	91.4
k-NN (k=7)	88.3	87.9	88.3	88.0

Random Forest achieved the highest accuracy (96.4%) with balanced precision and recall across all eight classes, attributable to its ensemble averaging across 200 trees and implicit feature selection through random subsampling at each split. ANN achieved the second-highest accuracy (94.8%), while SVM and k-NN were comparatively less effective, with k-NN in particular affected by the curse of dimensionality in the 42-dimensional feature space.

4.3 Noise Sensitivity

Table 2. Classifier accuracy under Gaussian measurement noise

Classifier	0% Noise	5% Noise	10% Noise	Degradation (pp)
Random Forest	96.4%	94.1%	91.8%	4.6
ANN (MLP)	94.8%	92.3%	89.5%	5.3
SVM (RBF)	91.7%	88.9%	85.2%	6.5
k-NN (k=7)	88.3%	84.6%	79.8%	8.5

All classifiers exhibited monotonic accuracy decline with increasing noise. Random Forest retained the highest absolute accuracy (91.8%) and smallest degradation (4.6 percentage points) at 10% noise, confirming suitability for field deployment where measurement noise is unavoidable; k-NN was the most noise-sensitive (8.5 percentage points), reflecting the general vulnerability of distance-based classifiers to local perturbation in feature space.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

4.4 Feature Category Contribution

Table 3. Random Forest accuracy by feature category

Feature Set	No. of Features	RF Test Accuracy (%)
Time-domain only	28	84.2
Frequency-domain only	9	87.6
Wavelet only	5	89.4
Combined (all)	42	96.4

Wavelet features achieved the highest standalone accuracy (89.4%) despite comprising only five features, indicating high information density. The combined multi-domain set outperformed the best individual category by 7.0 percentage points, confirming that time-, frequency- and wavelet-domain representations capture complementary, non-redundant information relevant to damage classification.

4.5 Confusion Matrix and Per-Class Behaviour

The healthy baseline (D0) was classified with 100% accuracy by the Random Forest classifier, indicating a very low false-negative risk at the basic detection level - the most safety-critical outcome for a practical SHM system. The lowest per-class accuracy (89.0%) was observed for D1, the lowest-severity single-element beam damage, most often confused with the healthy class, consistent with its small (1.6%) frequency shift. Classification confidence and per-class accuracy both increased monotonically with damage severity across all four classifiers, and multi-element damage cases (D6, D7) were classified with high accuracy (97-98%) owing to their larger overall stiffness reduction.

4.6 Feature Importance

Random Forest Gini-importance ranking identified the fundamental natural frequency (f1), mid-level wavelet energy coefficients (E3, E4) and mode shape curvature at Storey 2 as the most discriminative features, consistent with their established physical sensitivity to global and local stiffness change. Time-domain features, while individually ranked lower, contributed meaningfully to distinguishing damage cases with similar frequency signatures but different signal impulsiveness.

4.7 Sensor Configuration Effects

An ablation study showed that Sensor S3 (Storey 4) contributed most to overall accuracy, with its removal reducing accuracy by 4.3 percentage points, reflecting its proximity to the simulated damage locations and its role in capturing intermediate-height mode shape behaviour. This supports prioritizing sensor placement near anticipated damage zones and intermediate building heights rather than exclusively at the roof or base.

4.8 Comparison with Published Studies

Table 4. Comparison with representative published benchmarks

Study	Structure Type	Classifier	Reported Accuracy
Yan et al. (2007)	Numerical beam	SVM	~90%
Gonzalez & Zapico (2008)	Steel frame building	ANN	~88-93%
Gui et al. (2017)	Various	Optimized SVM	~92-95%
Avci et al. (2021), review range	Various (aggregate)	Multiple	85-98%
Present Study	G+5 RC frame building	Random Forest	96.4%

The 96.4% accuracy achieved in this study falls toward the upper end of the range reported in comparable published work, indicating that the developed framework performs competitively while extending prior research to the underrepresented case of multi-storey RC buildings evaluated with combined multi-domain features and explicit noise-robustness testing.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

5. CONCLUSIONS

5.1 Main Findings

Random Forest was the best-performing classifier across accuracy, noise robustness and per-class consistency. The combined multi-domain feature set substantially outperformed any individual feature category, confirming the complementary value of time-, frequency- and wavelet-based representations. The framework reliably detected early-stage (10% severity) single-element damage (89% accuracy) despite a corresponding frequency shift of only 1.6%, and classified the healthy baseline with 100% accuracy across all four classifiers.

5.2 Principal Numerical Findings

Key results are: 96.4% Random Forest baseline accuracy; 91.8% accuracy under 10% measurement noise; a 7.0 percentage-point improvement from multi-domain feature fusion over the best single category; natural frequency (f1) as the top-ranked discriminative feature; and Sensor S3 (Storey 4) as the single most informative sensor location (4.3 percentage-point contribution).

5.3 Practical Recommendations

Random Forest is recommended as the primary classifier for practical SHM deployment given its accuracy, noise robustness, computational efficiency (8.2 s training, 0.4 ms inference) and built-in feature-importance interpretability. Accelerometers should be prioritized at intermediate storey levels near likely damage zones. Given their strong standalone performance relative to dimensionality, wavelet features should be prioritized in resource-constrained onboard systems, while combined multi-domain extraction should be used wherever computational resources permit. Periodic model retraining and, for safety-critical structures, a dual-classifier (RF plus ANN) cross-validation approach are also recommended.

6. LIMITATIONS

All vibration data used in this study is numerically simulated; while the FE model was validated against empirical period estimates and standard modelling checks, simulated data cannot fully capture real-world variability from environmental effects, sensor installation imperfections or genuinely non-linear structural behaviour. Damage is idealized as a uniform stiffness reduction over the full element length rather than the spatially distributed, non-uniform nature of real cracking or corrosion. Environmental and operational variability (temperature, humidity, occupancy loading) are not explicitly modelled, representing an acknowledged limitation relative to field-deployed systems.

7. FUTURE SCOPE

Future work should prioritize experimental validation through scaled shake-table testing or deployment on a real instrumented building, extension to irregular, taller, or shear-wall/braced structural typologies, and investigation of deep learning architectures (1D-CNN, LSTM) given access to larger training datasets. Integration with real-time IoT-based wireless sensor networks and edge computing, explicit modelling of environmental and operational variability, transfer learning from simulated to real measured data, extension to continuous damage-severity regression, and unsupervised/semi-supervised approaches to reduce dependence on extensively labelled data are also identified as promising directions.

8. REFERENCES

[1] Rytter, A. (1993). Vibrational Based Inspection of Civil Engineering Structures. PhD Dissertation, Aalborg University, Denmark.
[2] Cawley, P. & Adams, R.D. (1979). The location of defects in structures from measurements of natural frequencies. Journal of Strain Analysis, 14(2), 49-57.
[3] 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.
[4] Salawu, O.S. (1997). Detection of structural damage through changes in frequency: A review. Engineering Structures, 19(9), 718-723.
[5] Pandey, A.K., Biswas, M. & Samman, M.M. (1991). Damage detection from changes in curvature mode shapes. Journal of Sound and Vibration, 145(2), 321-332.
[6] Stubbs, N., Kim, J.T. & Farrar, C.R. (1995). Field verification of a nondestructive damage localization and severity estimation algorithm. Proc. 13th Int. Modal Analysis Conf., 210-218.
[7] Fan, W. & Qiao, P. (2011). Vibration-based damage identification methods: A review and comparative study. Structural Health Monitoring, 10(1), 83-111.
[8] Carden, E.P. & Fanning, P. (2004). Vibration based condition monitoring: A review. Structural Health Monitoring, 3(4), 355-377.
[9] Worden, K., Manson, G. & Fieller, N. (1997). Damage detection using outlier analysis. Journal of Sound and Vibration, 229(3), 647-667.
[10] Sohn, H., Farrar, C.R., Hunter, N.F. & Worden, K. (2001). Structural health monitoring using statistical pattern recognition techniques. J. Dynamic Systems, Measurement, and Control, 123(4), 706-711.
[11] Yan, A.M., Kerschen, G., De Boe, P. & Golinval, J.C. (2007). Structural damage diagnosis under varying environmental conditions - Part I. Mechanical Systems and Signal Processing, 19(4), 847-864.
[12] Gonzalez, M.P. & Zapico, J.L. (2008). Seismic damage identification in buildings using neural networks and modal data. Computers & Structures, 86(3-5), 416-426.
[13] Rafiei, M.H. & Adeli, H. (2017). A novel machine learning-based algorithm to detect damage in high-rise building structures. The Structural Design of Tall and Special Buildings, 26(18).
[14] 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.
[15] Gui, G., Pan, H., Lin, Z., Yu, X. & Zhang, J. (2017). Data-driven support vector machine with optimization techniques for structural health monitoring and damage detection. KSCE J. Civil Engineering, 21(2), 523-534.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

[16] Santos, A., Figueiredo, E., Silva, M., Sales, C. & Costa, J. (2016). Machine learning algorithms for damage detection: Kernel-based approaches. Journal of Sound and Vibration, 363, 584-599.
[17] Staszewski, W.J. (1998). Structural and mechanical damage detection using wavelets. Shock and Vibration Digest, 30(6), 457-472.
[18] Sun, Z. & Chang, C.C. (2002). Structural damage assessment based on wavelet packet transform. Journal of Structural Engineering, 128(10), 1354-1361.
[19] Huang, N.E. et al. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. Royal Society A, 454(1971), 903-995.
[20] Nair, K.K., Kiremidjian, A.S. & Law, K.H. (2006). Time series-based damage detection and localization algorithm with application to the ASCE benchmark structure. Journal of Sound and Vibration, 291(1-2), 349-368.
[21] IS 1893 (Part 1):2016. Criteria for Earthquake Resistant Design of Structures. Bureau of Indian Standards, New Delhi.
[22] Chopra, A.K. (2012). Dynamics of Structures: Theory and Applications to Earthquake Engineering, 4th Ed. Pearson Education.
[23] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.
[24] Cortes, C. & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273-297.
[25] Lin, J.F. & Zhu, S. (2019). Application of artificial neural networks for inter-storey drift damage detection in steel frame buildings. J. Earthquake Engineering, 23(8), 1289-1308.
[26] Khoshnoudian, F., Esmaeili Niari, S. & Talaei, S. (2015). Damage detection of structures using frequency response function and neural network. Smart Structures and Systems, 16(2), 257-274.
[27] Sun, L., Wang, T. & Hu, X. (2020). Random forest-based seismic damage identification for a 12-storey reinforced concrete building. Structures, 28, 1851-1862.
[28] Bull, L.A., Cross, E.J., Worden, K. & Manson, G. (2018). A probabilistic framework for hierarchical classification in structural health monitoring. Mechanical Systems and Signal Processing, 109, 230-247.
[29] Abdeljaber, O., Avci, O., Kiranyaz, S., Gabbouj, M. & Inman, D.J. (2017). Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. Journal of Sound and Vibration, 388, 154-170.
[30] Peeters, B. & De Roeck, G. (2001). One-year monitoring of the Z24-Bridge: environmental effects versus damage events. Earthquake Engineering & Structural Dynamics, 30(2), 149-171.

SUPPLEMENTARY NUMERICAL SUMMARY

Table 5. Key numerical results

Metric	Result
Benchmark building	G+5 RC frame, 15 m x 15 m, 3 bays x 3 bays
Damage classes	8 (1 healthy + 7 damage scenarios)
Total samples / features	1600 samples / 42 features
Best classifier (clean data)	Random Forest - 96.4% accuracy
Best classifier (10% noise)	Random Forest - 91.8% accuracy
Weakest classifier	k-NN - 88.3% clean, 79.8% at 10% noise
Best individual feature category	Wavelet features - 89.4% standalone accuracy
Most discriminative feature	Fundamental natural frequency (f1)
Most challenging damage case	D1 - 10% beam stiffness reduction (89.0% RF accuracy)
Healthy baseline classification	100% accuracy, all classifiers
Most informative sensor	S3 - Storey 4 (4.3 pp contribution)
RF training / inference time	8.2 s / 0.4 ms per sample