Noise-Resilient Feature-Based Structural Health Monitoring Using Gradient Boosting

Ishan[*], Lakshmi Latha[†] and, Samit Ray-Chaudhuri[‡]

Abstract

This study presents a noise-resilient structural health monitoring (SHM) framework by using robust, multi-domain signal features and machine learning for accurate damage detection. A diverse set of features such as statistical, dynamic, frequency-domain, energy-based, higher-order, and complexity measures to capture diverse aspects of a signal's variability, structure, and frequency content are extracted and evaluated for robustness against noise to determine an optimal feature configuration for accurate detection of damage. Comparative analyses between baseline and optimized feature sets have been considered which demonstrated improved reliability under noisy conditions. The optimized features are then considerd on a 12-storey shear building model, where single-floor damage scenarios are simulated which is then used to validate the approach. A gradient boosting classifier is further trained and detailed floor-wise and feature-wise importance analyses are conducted to examine the spatial distribution of damage-relevant information. Additionally, the impact of noise on classification performance is also assessed, highlighting the significance of robust features in preserving model accuracy. The findings shows the potential of robust features in enhancing the resilience and interpretability of SHM systems under real-world noise conditions.

Keywords: Machine learning; Structural health monitoring; Robust features; 12-story building; Damage detection

[*]**Graduate Student, Department of Civil Engineering, IIT Kanpur, Kanpur, UP-208016**

[†]Post-Doctoral Researcher, Department of Civil Engineering, IIT Kanpur, Kanpur, India-208016

[‡]Professor, Department of Civil Engineering, IIT Kanpur, Kanpur, India-208016, Email: samitrc@iitk.ac.in, Tel.: +91 512 2597267 (Corresponding Author)

1 Introduction

Structural health monitoring systems have been utilized in various field and civil engineering is one such field for which SHM has become really important. It is because of the simple reason of the need to understand, preserve, and extend the life of built structures [1]. Initially, building monitoring was related to understanding performance during earthquakes and storms. For that low-amplitude vibration testing and then long-term monitoring of real loading events was considered. Programs like California Strong Motion Instrumentation Program is aimed to improve structural behaviour under earthquake by collecting ground motion data. In high-rise buildings, monitoring considered wind-induced responses and occupants comfort by ensuring that response are within safe and comfortable limits [2]. Modern SHM involves using high-quality accelerometers and data acquisition tools. This along with random decrement techniques for tracking fundamental mode damping and building tilt has been considered in tall buildings of Hong Kong [2]. Today, SHM increasingly relies on database management, system identification, and artificial intelligence which require sophisticated sensors, real-time data acquisition, and advanced analytical capabilities.

Structural health monitoring as such is the process of executing a damage identification strategy by continuously observing it over time by using sensors (e.g., strain gauges, accelerometers, acoustic emission sensors) and data acquisition systems. It further involves collecting measurements at regular intervals, the extraction of appropriate damage-sensitive features from these measurements and the subsequent analysis of these features to determine systems current health [3]. If there is a structural damage,it usually causes a decrease in structural stiffness and an increase in structural damping thus changing vibration characteristics. Vibration data widely used to detect structural damage includes modal data and the frequency response function [4, 5]. Modal data includes the natural frequency, mode shape, mode shape curvature, modal flexibility, and modal strain energy [6, 7]. But the challenge is that vibration response from civil engineering structures may exhibit variations over both time and space. The data may be exposed to several uncertainties or contamination, which can pose challenge during damage identification from vibration data. One such issue is the presence of noise in the collected data. Noise may be due to environmental factors, problems with sensors, or due to errors in data transmission. Hence, denoising techniques were greatly explored with an aim to improve the

quality of structural vibration data. Traditional noise filtering techniques based on fourier analysis may not achieve effective removal of noise when dealing with signal and spectra overlap. This is due to the fact of removal of essential signal components by the filter in place of noise. While wavelet-based denoising is a time-frequency analysis method of signal that allows for multi-resolution analysis. This preprocessing method can help in the gain of performance under noisy environment [8, 9]. In this, the signal components are decomposed into frequency bands with time localization making the approach more effective, robust and superior to Fourier analysis algorithm [10]. This will improve the accuracy of damage detection, condition assessment, and also predictive maintenance [11, 12]. Feature extraction is an important stage in signal analysis which reduces the dimensionality of signals by capturing key characteristics essential for subsequent tasks analysis [13, 14]. So, there is need for extraction of features from the signal that reflect underlying changes to analyze behaviour of any structure through monitoring [15]. Hence, from the denoised signal for getting more important information multi-domain feature extraction derived from time-domain, frequency-domain, and time-frequency domain analyses is needed [16, 17]. This approach enables the capture of transient, spectral, and statistical attributes, which can enhance the accuracy and robustness of structural behavior characterization both globally and locally. Even integrating multi-domain features have also observed class separability and reduced classification errors [18, 19]. Limiting analysis to a single domain may not give an effective criterion. The performance may depend on recognition rate, signal length, response time, computational resources, and complexity of the method as well [20]. It may not necessarily provide noise-resilient features but the performance of approach may be good with minimum value for the criterion.

Hence, while multi-domain feature extraction ensures capturing of wide range of signal information, robust features need to be extracted from various features to actually to remain effective with the real-world, noise-prone environments. Robust features, it will further ensure accurate and reliable interpretation of sensor data during structural health condition assessment. Even under noise from environmental and operational sources, robust features provide stable inputs for model updating and analysis, allow meaningful trend analysis despite varying conditions and increase effectiveness of SHM. Even with robust features in place, SHM techniques which rely heavily on physics-based models or threshold-based rules, are limited in their ability to capture complex, nonlinear damage patterns making it less adaptable to chang-

ing structural conditions. Machine Learning (ML) can significantly enhances the capabilities of SHM systems making them more effective than traditional approaches [21–23]. ML models can automatically learn from robust features, recognize hidden patterns, and manage large datasets easily. They also enable advanced tasks like feature importance analysis, anomaly detection, and predicting future damage [24]. ML supports real-time, scalable monitoring on large infrastructure networks that traditional methods cannot achieve effectively. Thus, with improved data representation using robust features ML can deliver adaptive, and efficient solutions making the best use of that data in SHM. Such data-driven SHM strategies is most applicable to tall buildings as well, which is one of the complex civil engineering structure due to its height, design and exposure to various environmental and operational loads. These structures exhibit highly varying non-linear responses throughout their heights. In such cases, reliable classification of states of structure is essential across different floor levels. Robust feature extraction methods along with ML models, damage will be easier to detect which will make the interpretation of structural behaviour more accurately. Such methods can give more clear behaviour making sensor deployment in a more targeted manner. This can also give deeper insights regarding structure contributing to accurate classification [25]. This approach can capture damage-sensitive information that varies significantly across floors using machine learning systematically.

This study presents the development and evaluation of a noise-resilient structural health monitoring approach using robust, multi-domain features from signals and machine learning techniques. A comprehensive set of features is extracted from signals such as statistical, spectral, autocorrelation, energy-based, and fractal features. The robustness of these features is considered to evaluate the optimal feature configuration for damage-sensitive representation. A comparison between baseline and optimized configurations has been considered. This comparison is to highlight the improvements in feature reliability under noise-prone conditions while using robust features. Using the robust features,a study is conducted on a 12-storey shear building model where in each case only one specific floor is damaged and the rest are considered undamaged. A Gradient Boosting Classifier is trained and used to identify damage patterns across floors. Floor-wise and feature-wise importance analyses are conducted to determine the spatial distribution regarding damage-relevant information. Further, the study examines the effect of noise on model performance, assesses feature importance, and further explores

the correlation between floor-level feature significance and classification accuracy. This study employs noise-robust multi-domain feature optimization, precise floor-level damage localization, and interpretable gradient boosting classifiers, by providing enhanced robustness, spatial resolution, and explainability as compared to conventional methods that typically utilize single-domain features, assume noise-free data, and lack detailed interpretability or localized damage identification.

2 Methodology for Noise-Robust Feature Extraction

Structural Health Monitoring (SHM) primarily depends on sensors for condition assessment of structures. During damage or in any unusual changes, these sensors gives critical information valuable for assessment. But, the sensor data may be often contaminated with noise that can lead to inaccurate assessments. This research work focusses on a comprehensive methodology, which focusses on robust feature extraction from the signals against noise interference. This is aimed at getting the right information even if the signal is noisy.

To simulate real-world measurement scenarios, noise-free sinusoidal signal is considered in this study which serves as a baseline comparison. The sinusoidal signal generated on the central processing unit (CPU) is given by,

$$x(t) = \sin(0.01\pi t), \quad t \in [0, 1000], \quad \text{with 100,000 samples} \tag{1}$$

where, $x(t)$ is clean sinusoidal signal as a function of time, t is time variable, where $t \in [0, 1000]$ and π is approximately 3.14.

For efficient high-dimensional processing, the signal was then transferred to graphics processing unit (GPU) memory using GPU accelerated Python library. To simulate real-world conditions, Gaussian noise is used to model random noise. The noise is scaled by a factor proportional to the standard deviation of the original signal for various levels. Noise added in the form of scaled Gaussian white noise is given by,

$$x_{\text{noisy}} = x(t) + \mathcal{N}(0, \sigma^2), \quad \text{where } \sigma = \text{noise\_factor} \cdot \text{std}(x) \tag{2}$$

where, x_{noisy} is noisy version of the signal, $\mathcal{N}(0, \sigma^2)$ is Gaussian (normal) noise with

mean 0 and variance σ^2, σ is standard deviation of the Gaussian noise, computed as $\sigma = \text{noise\_factor} \cdot \text{std}(x)$, noise_factor is a scalar multiplier that adjusts the strength of the added noise and std(x) is standard deviation of the clean signal $x(t)$. Noise levels has been varied from 1% to 30% (linearly spaced over 5 points) to simulate increasing sensor uncertainty.

Further,wavelet-based denoising technique is used to filter out noise. The thresholding method using a discrete wavelet transform (DWT) was considered.This will preserve the important signal which retains underlying structural characteristics while removing noise which are typically small in value [11]. The noisy signal data can be decomposed into wavelet coefficients of one set or multiple sets of detail coefficients at various scales. The noise standard deviation is further estimated from the median of the absolute values of detail coefficients. Further, the threshold is calculated based on noise estimate and segment length [11]. The threshold may be applied to the detail coefficients by using soft thresholding (i.e. shrinks coefficients toward zero, reducing the magnitude of small coefficients while preserving continuity) or hard thresholding (i.e. sets coefficients with magnitude less than λ to zero, completely removing minor components). After thresholding, reconstruction of denoised signal is considered with the modified coefficients using the inverse wavelet transform.

The denoised signal is further used to extract input features and further, the robust features.

2.1 Input Features

To ensure a comprehensive and noise-resilient understanding of signal behavior, a diverse set of 50 features is extracted. Ranging from basic statistical measures (like mean, median, and standard deviation), to dynamic descriptors (such as velocity and jerk), frequency-domain indicators (like Fast Fourier Transform (FFT) amplitude and power spectral density), energy-based metrics (including spectral energy and Teager-Kaiser Energy Operator (TKEO)), higher-order statistics (skewness, kurtosis, entropy), and complexity measures (such as fractal dimension and non-zero count) with each of the feature capturing a unique aspect of the signal's variability, structure and frequency content is further considered.

2.1.1 Basic Statistical Features

A set of single-number statistics or descriptive statistics can be used to understand central tendency, spread ,variability and nature of distribution of dataset [26]. Mean, median, standard deviation, minimum and maximum values of dataset are basic statistical features used for dataset in this research and are discussed below. Mean and meadian are measures of location which can provide some quantitative values regarding the center (or central tendency), or some other location of the dataset (or spread of the signal) . The mean of dataset is given by,

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3}$$

where, $\bar{x}$ is the arithmetic average of the dataset values, x_i is i-th value in the dataset , and n gives the number regarding data points.

Median is reflection of central tendency of given sample and is insensitive to extreme values or outliers. The median of dataset is given by,

$$\tilde{x} = \begin{cases} x_{(n+1)/2}, & \text{if } n \text{ is odd} \\ \frac{1}{2}\left(x_{n/2} + x_{n/2+1}\right), & \text{if } n \text{ is even} \end{cases} \tag{4}$$

where, $\tilde{x}$ is the median of the dataset values (provided that the values in the dataset are arranged in the increasing order of magnitude), x_n denotes the n-th after sorting the dataset in ascending order.

Standard deviation gives measure of spread of the dataset around its mean μ, indicating the variability of given signal. The deviation of dataset is given by,

$$\mathrm{s} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \mu\right)^2}$$

where, s is the standard deviation of the dataset, x_i gives the i-th positions data point, μ is the mean of given dataset, and n gives the total number of observations.

The minimum and maximum values of the signal are the lowest and highest points in the dataset and are given by ,

$$x_{\min} = \min\{x_1, x_2, \ldots, x_n\} \tag{5}$$

and

$$x_{\max} = \max\{x_1, x_2, \ldots, x_n\} \tag{6}$$

where, $x_1, x_2, ..., x_n$ are the dataset values.

2.1.2 Velocity and Jerk Features

Velocity and jerk features are based on the rate of change of the signal (velocity) and the rate of change of velocity (jerk).

Velocity is the first derivative of the signal with respect to time and is given by,

$$v(t) = \frac{1}{\Delta t}\left(x(t+1) - x(t)\right) \tag{7}$$

where, $v(t)$ is velocity at time step t, $x(t)$ is position or displacement at time step t and Δt is time interval between successive samples.

Mean absolute velocity provides an overall measure of the speed and is given by,

$$\text{Mean absolute velocity} = \frac{1}{n-1}\sum_{i=1}^{n-1}\left|\frac{x_{i+1} - x_i}{\Delta t}\right| \tag{8}$$

where, x_i is position or signal value at time step i, Δt is time interval between successive samples and n is total number of samples.

Jerk is the rate of change of velocity, or the second derivative of the signal. Jerk quantifies the smoothness or abruptness of changes in motion.

$$j(t) = \frac{1}{\Delta t}\left(v(t+1) - v(t)\right) \tag{9}$$

where, $j(t)$ is jerk at time step t, $v(t)$ is velocity at time step t and Δt is time interval between successive samples.

Mean absolute jerk can capture overall smoothness or irregularity in acceleration and is

given by,

$$\text{Mean absolute jerk} = \frac{1}{n-2}\sum_{i=1}^{n-2}\left|\frac{v_{i+1}-v_i}{\Delta t}\right| \tag{10}$$

where, v_i is velocity at the i-th time step, Δt is time interval between successive samples and n is total number of samples.

2.1.3 Zero Crossing Rate and Related Features

Zero crossing rate is computed by counting the number of times the signal changes its sign and is given by,

$$\begin{aligned} Z(i) &= \frac{1}{2W_L}\sum_{n=1}^{W_L}\left|\operatorname{sgn}\left[x_i(n)\right]-\operatorname{sgn}\left[x_i(n-1)\right]\right|, \\ &\text{where } \operatorname{sgn}() \text{is the sign function, i.e.} \\ \operatorname{sgn}\left[x_i(n)\right] &= \begin{matrix} 1, x_i(n)\geq 0, \\ -1, x_i(n)<0. \end{matrix} \end{aligned} \tag{11}$$

Sign changes are the total mumber of sign changes in the signal and is given by,

$$\text{Sign changes} = \sum_{i=1}^{n-1}\mathbf{1}\left[x_i \cdot x_{i+1} < 0\right] \tag{12}$$

where, x_i is signal value at the i-th time step, n is total number of samples, 1 is indicator function that equals 1 if the condition is true, and 0 otherwise.

2.1.4 Autocorrelation Features

Autocorrelation measures the similarity between the signal and a shifted version of itself, which helps identify periodic patterns. Lag1_autocorrelation is the autocorrelation at lag 1 (shifted by one time step) and is given by,

$$\text{Lag1\_autocorrelation} = \frac{\sum_{i=1}^{n-1}\left(x_i-\mu\right)\left(x_{i+1}-\mu\right)}{\sqrt{\sum_{i=1}^{n-1}\left(x_i-\mu\right)^2 \cdot \sum_{i=1}^{n-1}\left(x_{i+1}-\mu\right)^2}}$$

where, x_i is the value of the signal at time step i, μ is the mean of the entire signal given by, $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$ and n is the total number of observations.

Lag2_autocorrelation is the autocorrelation at lag 2 (shifted by two timesteps) is given by :

$$\text{Lag2\_autocorrelation} = \frac{\sum_{i=1}^{n-2} (x_i - \mu)(x_{i+2} - \mu)}{\sqrt{\sum_{i=1}^{n-2} (x_i - \mu)^2 \cdot \sum_{i=1}^{n-2} (x_{i+2} - \mu)^2}}$$

where, x_i is the value of the signal at time step i, μ is the mean of the entire signal given by, $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$ and n is the total number of observations.

2.1.5 Spectral Features

Spectral features are derived from the frequency domain representation of the signal using techniques like the Power Spectral Density (PSD) and Fast Fourier Transform (FFT). The Power Spectral Density (PSD) represents how power is distributed across frequencies, and is calculated using methods such as Welch's method:

$$\text{PSD}(f) = |\mathcal{F}(x(t))|^2$$

where, $\text{PSD}(f)$ denotes power spectral density at frequency f, $x(t)$ is the signal as a function of time and $\mathcal{F}(x(t))$ is the Fourier transform of the signal.

The Fast Fourier Transform (FFT) provides the frequency content of the signal and is expressed as ,

$$\text{FFT Amplitude} = |\mathcal{F}(x(t))|$$

where, $x(t)$ is time-domain signal as a function of time t and $|\mathcal{F}(x(t))|$ is magnitude of the Fourier Transform, i.e., amplitude spectrum. This magnitude represents the strength of different frequency components.

2.1.6 Wavelet Transform Features

Wavelet transforms analyze the signal at different scales, capturing both time and frequency features. For signal $x(t)$, the Continuous Wavelet Transform (CWT) at scale s is given by:

$$\text{CWT}(s,t) = \int_{-\infty}^{\infty} x(t')\, \psi\left(\frac{t - t'}{s}\right) dt'$$

where ψ is the wavelet function, and s is the scale.

2.1.7 Energy-Based Features

Energy-based features quantify the signal's energy distribution and overall strength. Spectral energy is the total energy in the signal which is the sum of the squares of the signal values and is given by,

$$\text{Spectral energy} = \sum_{i=1}^{n} x_i^2$$

where, x_i is the i-th spectral component (e.g., FFT magnitude or signal amplitude at frequency bin i) and n is total number of components or frequency bins.

Teager-Kaiser Energy Operator(TKEO) measures signal energy, emphasizing high-frequency components and is given by,

$$\text{TKEO } = x_{i+1}^2 - x_i - x_{i+2}$$

where, x_i is signal value at the discrete time index i, x_{i+1} is signal value at the next time index $i+1$ and x_{i+2} is signal value at the next time index $i+2$.

2.1.8 Higher-Order Statistical Features

These features can capture higher-order statistical moments of the signal, such as skewness and kurtosis. Skewness measures the asymmetry of the signal's distribution and is given by,

$$\text{Skewness}(x) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma} \right)^3$$

where, x_i is i-th data point in the dataset, μ is mean of the dataset, which is given by $\mu = \frac{1}{n} \sum_{i=1}^{n} x_i$, σ is standard deviation of the dataset, given by $\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2}$ and n is total number of data points.

Kurtosis measures the tailedness of the distribution and is given by,

$$\text{Kurtosis}(x) = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$$

where, $\mathrm{Kurtosis}(x)$ is sample kurtosis, adjusted for small sample bias, x_i is i-th data point in the sample, μ is sample mean, i.e., $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$, σ is sample standard deviation, i.e., $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2}$ and n is number of data points.

Entropy quantifies the uncertainty or randomness in the signal and is given by,

$$\text{Entropy}(x) = -\sum_{i=1}^{n} p(x_i)\log\left(p(x_i)\right)$$

where, x_i is i-th possible value of the signal or random variable x and $p(x_i)$ is the probability of occurrence of x_i and n is the total number of discrete outcomes.

2.1.9 Fractal and Other Features

These features quantify the complexity and structure of the signal, providing insights into its self-similarity and irregularity. Fractal dimension quantifies how the signal's structure scales with a change in scale and is given by,

$$\text{Fractal dimension } = \log\left(N_{\text{boa}}(\epsilon)\right) / \log(1/\epsilon)$$

where,$N_{\text{box}}(\epsilon)$ is the number of boxes of size ϵ needed to cover the signal or structure and ϵ is the size (scale) of each box.

Non-zero count is the number of non-zero elements in the signal and is given by,

$$\text{Non-zero count} = \sum_{i=1}^{n} \mathbb{I}(x_i \neq 0)$$

where, x_i is the i-th value in the signal or vector, n is the total number of elements, and $\mathbb{I}(x_i \neq 0)$ is the indicator function, which equals 1 if $x_i \neq 0$, and 0 otherwise.

Further, noise robustness is considered by optimization by grid search where different combinations of setups which includes wavelet family type, wavelet level, thresholding mode, sign threshold etc.

2.2 Parameter Grid Search

A comprehensive parameter grid search was aimed at determining the optimal configuration for wavelet-based signal feature extraction. The goal was to minimize the average slope of the extracted features, thereby ensuring robustness such that it barely changes with noise in the signal representation.

The grid search spanned four primary parameters which includes wavelet family, decomposition level, thresholding mode, and zero-crossing threshold and details are given in Table 1. Before feature extraction from a noisy signal, two wavelet families of Daubechies 4 (db4) and Symlet 4 (sym4) is considered, which basically defines the shape of basis functions that will be used to break down the signal.

The decomposition level specifies how many times the signal is broken down into low and high frequency components. Levels 1, 2, and 3 were tested to examine signal detail at multiple scales. The thresholding mode, either soft or hard, controls how noise is removed while wavelet detail coefficients gets modified. Soft thresholding shrinks all coefficient values toward zero, while hard thresholding zeroes out values below a threshold and keeps the rest unchanged. Zero-crossing threshold was considered ignoring very small changes in the signal due to noise. Only zero-crossings with a surrounding minimum magnitude difference greater than 0.0, 0.01, or 0.02 were counted, to make the features more stable.

Table 1: Parameter grid for wavelet-based feature extraction

Parameter	**Values Explored**
Wavelet Family	['db4', 'sym4']
Decomposition Level	[1, 2, 3]
Thresholding Mode	['soft', 'hard']
Zero-Cross Threshold	[0.0, 0.01, 0.02]

Each unique combination of parameters from the grid in Table 1, average slope of the extracted features were considered. Then optimization of slope metric, which is a quantitative measure of feature variability, is further considered. For each setup, noise levels are considered which varied from 1% to 30%. With the change of noise, average slope of the value change for each feature is calculated. If the feature stays same with flat slope, it may be categorized as robust. The best set-up in this case may be the one which shows very little or no change

with variation of noise levels. Sensitivity of feature to increasing noise was quantified with robustness metric. If f represent the value of a particular feature, and $\eta \in [0.01, 0.3]$ denote the noise level (which ranges from 1% to 30%). The robustness score is given by,

$$\text{Robustness}(f) = \left| \frac{df}{d\eta} \right| \tag{13}$$

where, Robustness(f) represents the robustness score of the feature f, which quantifies the feature's sensitivity to noise, variable f denotes the numerical value of a specific extracted feature, such as mean, energy, entropy, or skewness and η refers to the noise level, typically expressed as a fraction of the signal's standard deviation, ranging from 0.01 to 0.3 (i.e., 1% to 30%). The derivative $\frac{df}{d\eta}$ gives the rate at which the feature value changes as the noise level increases. Taking the absolute value ensures only the magnitude of the change is considered, regardless of direction. Lower robustness score indicates that the feature f is less sensitive to changes in noise level, and hence more stable. During parameter optimization of robust features, this robustness score served as the objective function. The main goal was to identify the parameter set that minimizes the average slope across all features:

$$\text{Minimize: } \frac{1}{N} \sum_{i=1}^{N} \left| \frac{df_i}{d\eta} \right| \tag{14}$$

where N, is the total number of features. This ensures selection of parameters that yield features with high resistance to noise.

2.3 Best Parameter Configuration

After all tested parameter configurations, the one which yielded the most noise-robust features was observed as: wavelet family = sym4, wavelet level = 3, wavelet mode = soft, and sign threshold = 0.02. This combination exhibited lowest average slope metric with varying noise levels. The structure of the signal was preserved with the use of sym4 wavelet, which have a nearly symmetric shape. Level 3 decomposition allowed for a deeper analysis to capture broader signal patterns while not using much details. Using level 3 decomposition allowed the method to capture broader patterns in the signal without losing too much detail. The use of soft thresholding could provide a smoother attenuation of noise, avoiding abrupt changes

that can distort underlying signal features. The nonzero sign threshold helped to prevent trivial sign changes that may be caused by minor fluctuations in noisy signals. Hence, this parameter configuration was found to be robust making it suitable for real-world applications involving unpredictable and varying noise conditions.

2.4 Comparison of Baseline vs. Optimized Configuration

To evaluate the effectiveness of the optimized parameter configuration, feature trajectories were analyzed under varying noise levels and compared against the baseline configuration. The baseline configuration consisted of wavelet family = db4, wavelet level = 1, wavelet mode = soft, and sign threshold = 0.0. In contrast, the optimized configuration employed the parameters sym4, level 3, soft thresholding, and a sign threshold of 0.02. Feature values were plotted as functions of increasing noise to demonstrate their stability under both configurations. Figures 1–6 present the visual results across all configurations. As shown in Figures 1–6, the optimized configuration demonstrated flatter and smoother feature trajectories across all tested noise levels. This indicates a significant reduction in sensitivity to noise, which validates the robustness of the selected parameters. The comparison of average robustness scores between baseline and optimal configurations shows that the baseline setup achieved an average robustness score of 0.0372. In contrast, the optimal configuration achieved a significantly lower average robustness score of 0.0211, representing a 43.3% improvement in robustness.

The optimal configuration demonstrated reduced sensitivity across nearly all feature categories. In particular, entropy and higher-order statistical features exhibited noticeably flatter trajectories under increasing noise, indicating enhanced resilience to sensor degradation and improved generalizability in practical SHM deployments.

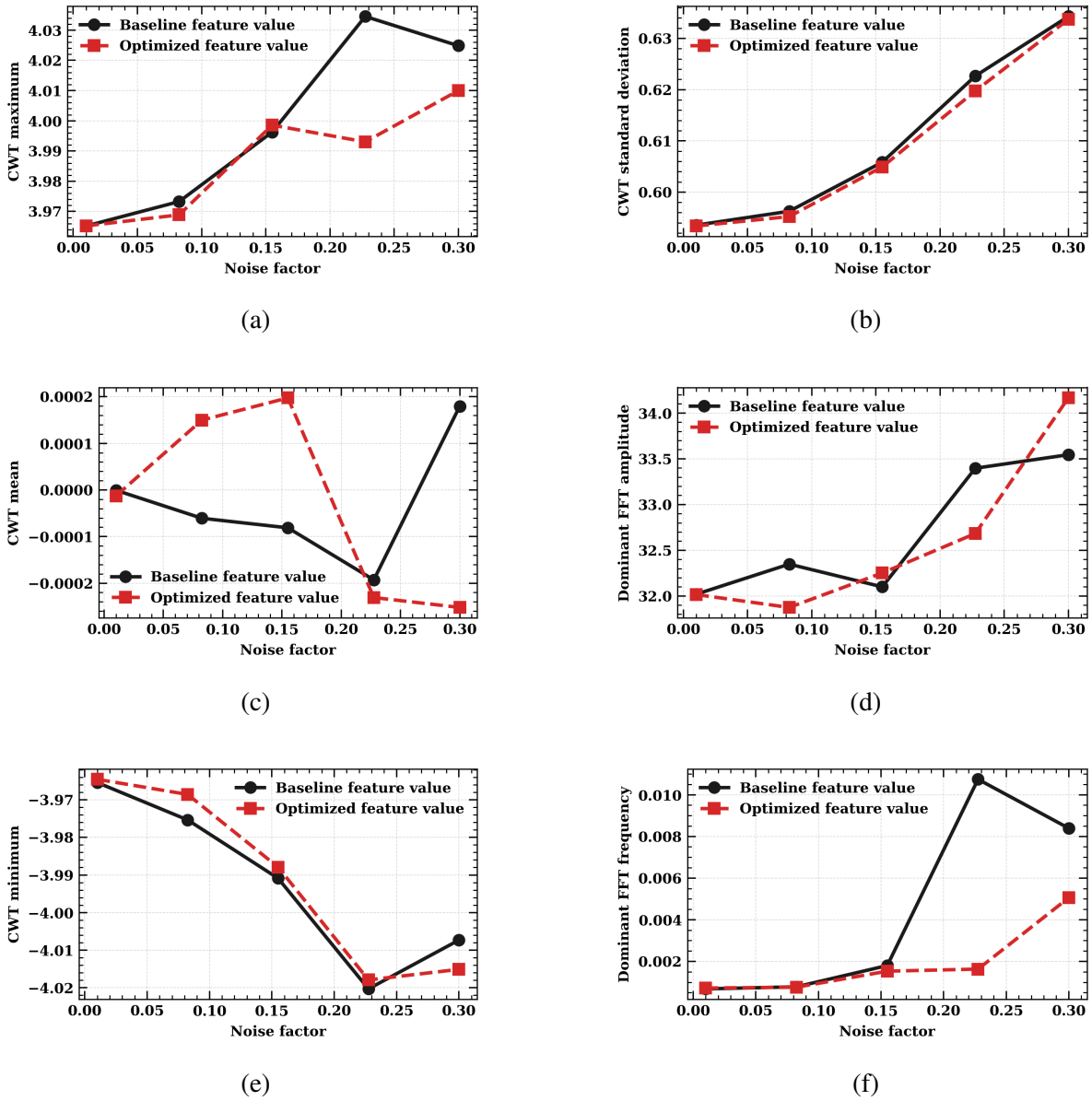


Figure 1: Baseline and optimized feature values across varying noise factors for: (a) Continuous wavelet transform maximum, (b) Continuous wavelet transform standard deviation, (c) Continuous wavelet transform mean, (d) Dominant FFT amplitude, (e) Continuous wavelet transform minimum, (f) Dominant FFT frequency.

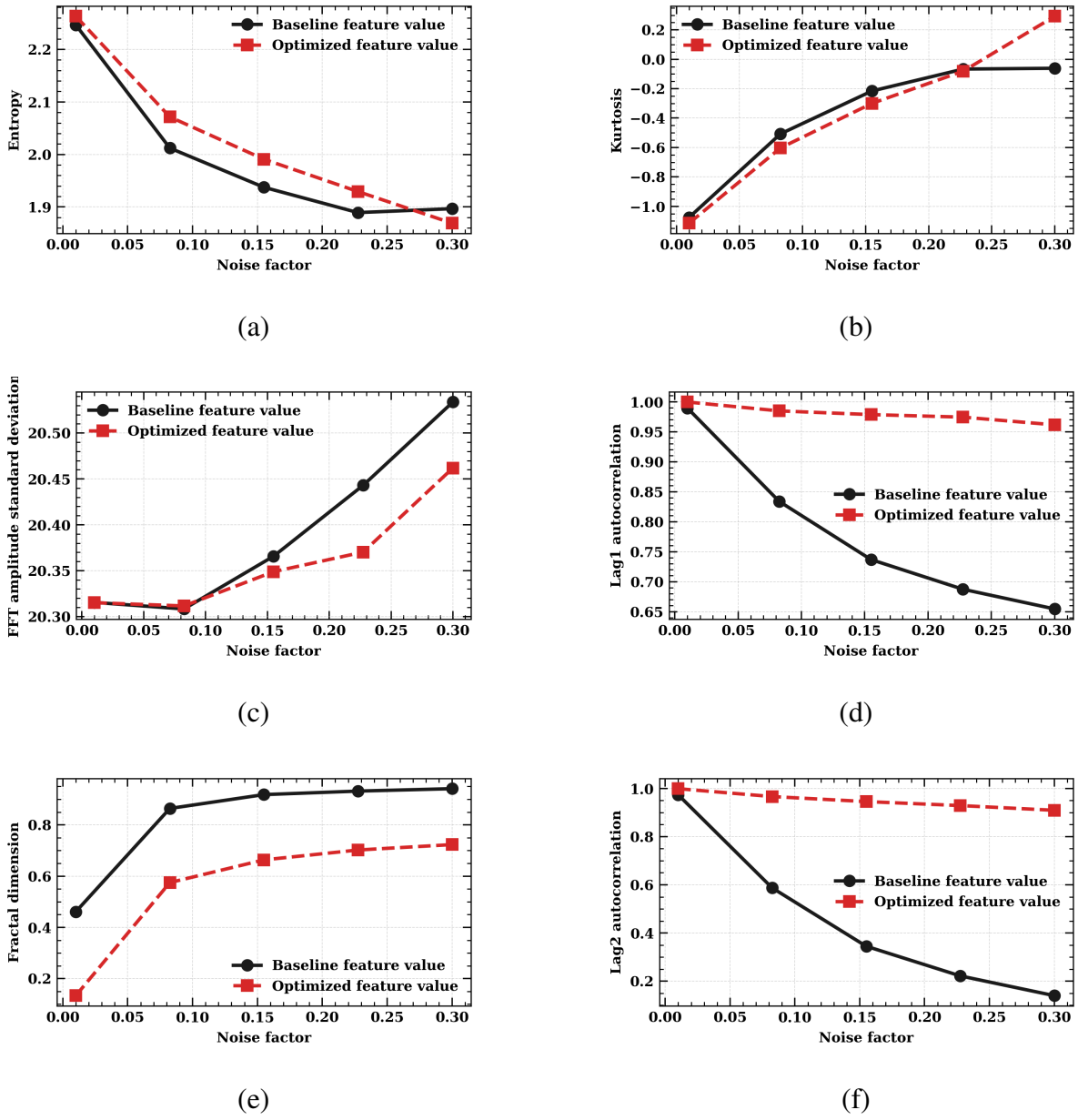


Figure 2: Baseline and optimized feature values across varying noise factors for: (a) Entropy, (b) Kurtosis, (c) FFT amplitude standard deviation , (d) Lag1 autocorrelation, (e) Fractal dimension , and (f) Lag2 autocorrelation.

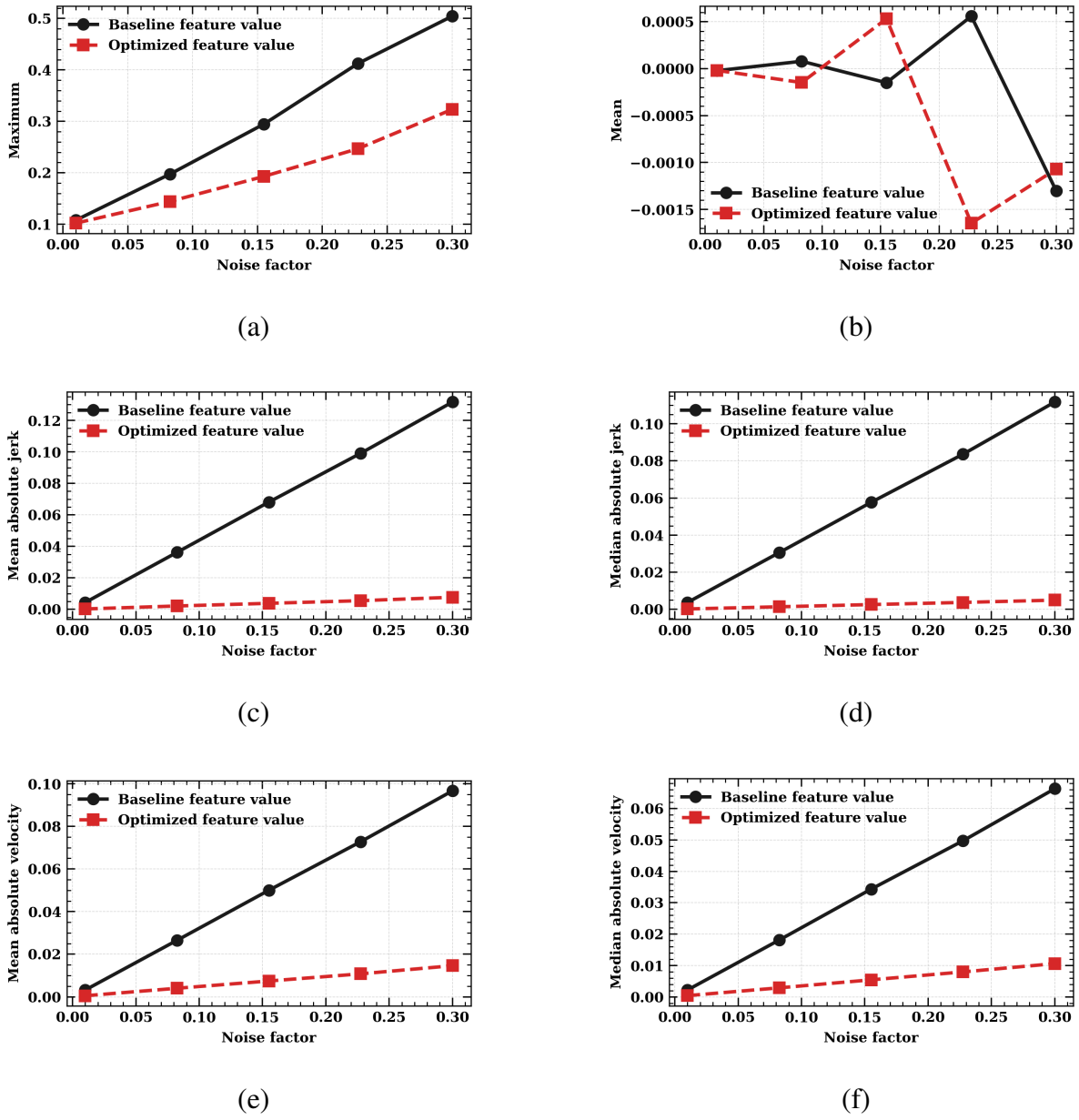


Figure 3: Baseline feature value and optimized feature value for varying noise factors: (a) Maximum, (b) Mean , (c) Mean absolute jerk , (d) Median absolute jerk , (e) Mean absolute velocity, and (f) Median absolute velocity.

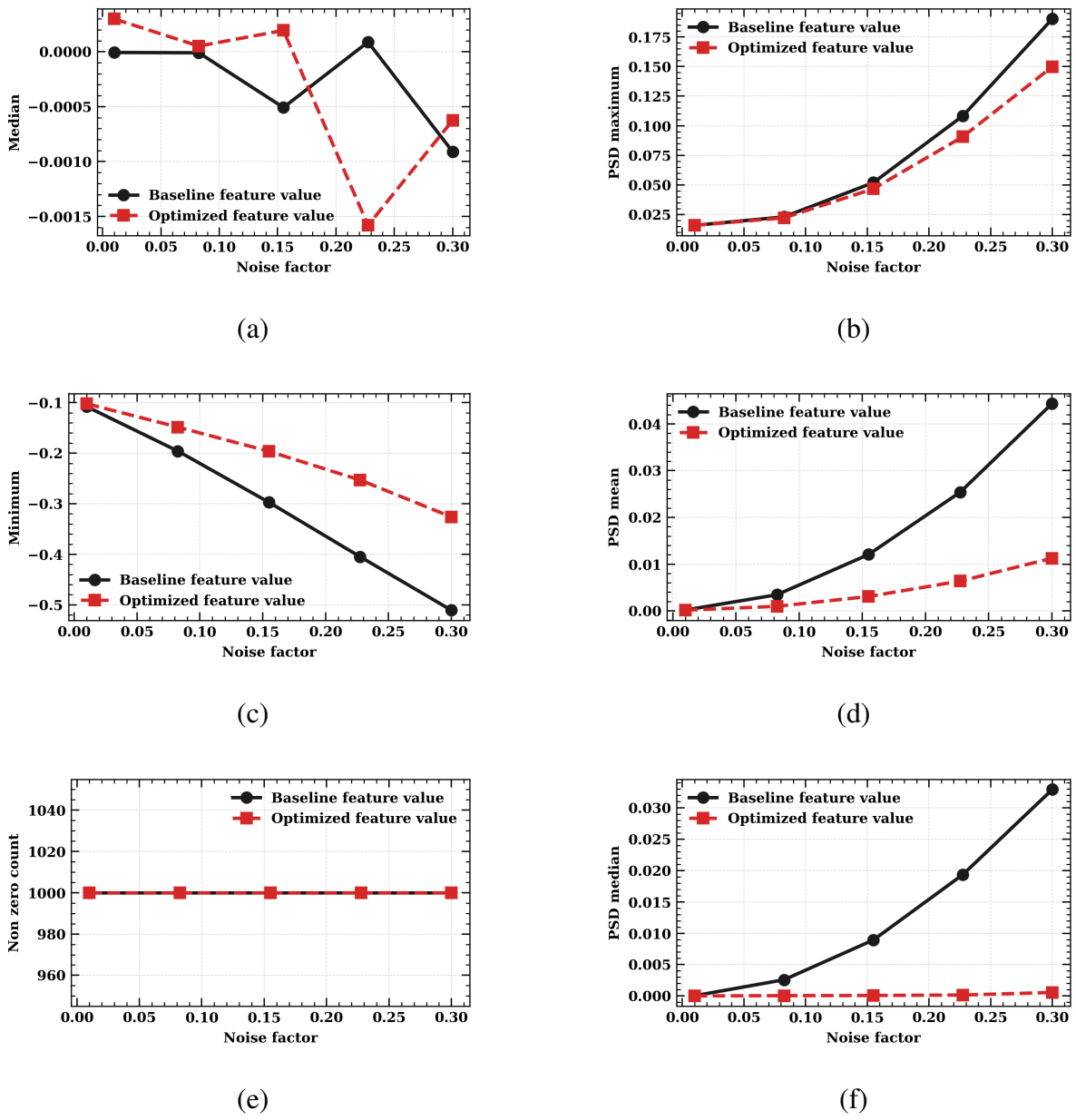


Figure 4: Baseline feature value and optimized feature value for varying noise factors: (a) Median, (b) PSD maximum, (c) Minimum, (d) PSD mean, (e) Non-zero count, (f) PSD median.

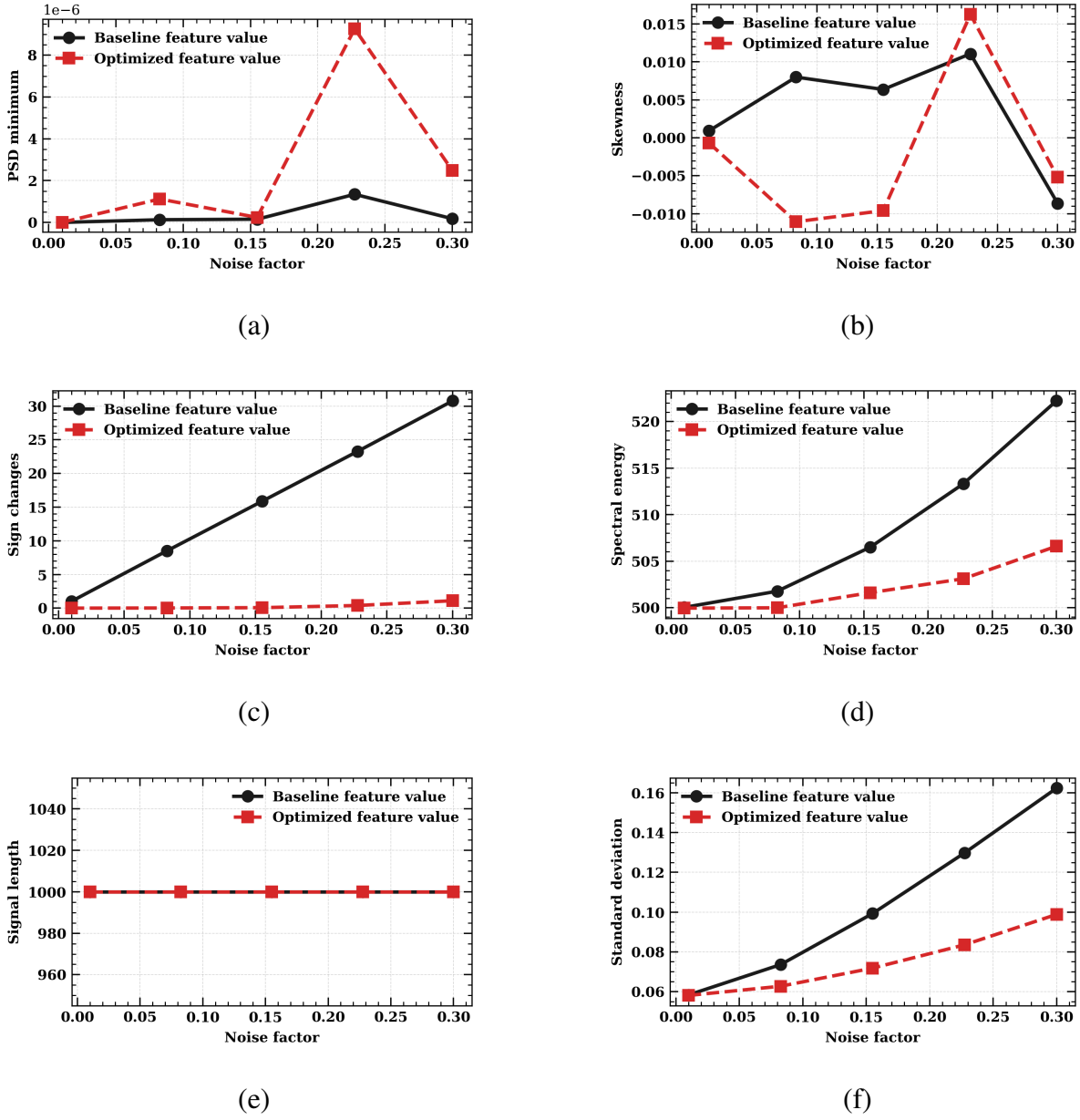
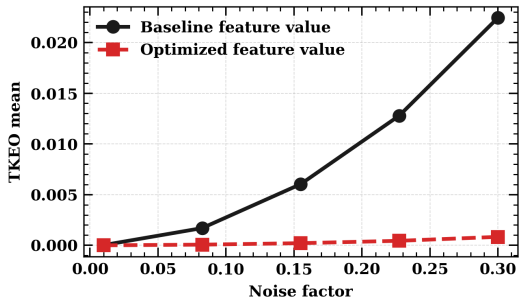
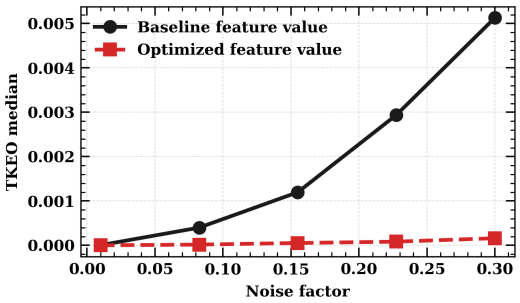


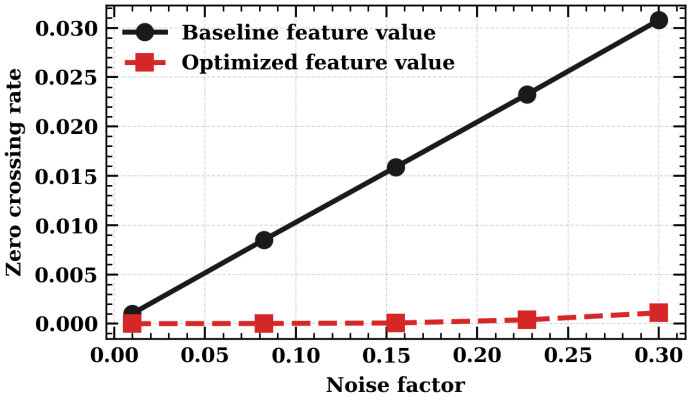
Figure 5: Baseline feature value and optimized feature value for varying noise factors: (a) PSD minimum, (b) Skewness, (c) Sign changes, (d) Spectral energy, (e) Signal length, (f) Standard deviation.



(a)



(b)



(c)

Figure 6: Baseline feature value and optimized feature value for varying noise factors: (a) TKEO mean, (b) TKEO median, (c) Zero crossing rate.

3 Case study of 12-story shear building

The study considered a 12-story shear building which is modelled using OpenSees version 2.4.1 [27] subjected to nonlinear dynamic analysis. The building is simplified as a lumped-mass shear building model. The base is fixed in all directions. Each floor is assigned a uniform lumped mass of 1.0e5 kg, capturing the inertial effects. The interstory stiffness is modeled using truss elements connecting adjacent nodes. These elements simulate the lateral stiffness between floors. The material for the truss elements is elastic with Youngs Modulus, E of 2.00e8 N/m², and each has the same cross-sectional area, ensuring uniform stiffness between stories. Rayleigh damping with 5% critical damping is included for the first two natural modes. The damping coefficients are automatically computed from the eigen frequencies of the system.

Dynamic loading is applied in the form of a uniform base excitation and transient analysis is carried out using the Newmark integration method.

3.1 Floor-wise and Feature-wise Importance Analysis

Structural acceleration response corresponding to various damage levels in the 12-story building has been considered in this study. Each floor has been defined with six damage levels ranging from 0.05 to 0.30 and maps each level to a unique integer label from 1 to 6.

Inorder to facilitate the classification of structural damage based on the data, a robust feature extraction and data processing pipeline considered. The acceleration data with defined damage levels for 12 floors are considered for feature extraction. For that, the signals are segmented into non-overlapping windows of 256 samples, and from each segment, eight robust time-frequency domain features are extracted. These features include signal length, non-zero count, Teager-Kaiser Energy Operator (TKEO) mean and median, power spectral density (PSD) mean and median, sign changes, and zero-crossing rate. Each feature is computed per signal channel resulting in 96 features (8 features per floor) and their corresponding labels with respect to the damage. This labeled dataset is further considered for machine learning algorithm of gradient boosting classifier for damage detection or classification. For that feature values were standardized using z-score normalization prior to training. Each model was trained on 80% of the data and evaluated on the remaining 20%.

Feature importance calculated using the gradient boosting classifier (GBC) based on mean decrease in impurity (MDI) and is given in Figure 7. The basis of MDI is based on how often each feature is used effectively to split the data in the decision tress that will result in the machine learning model. The classification error or impurity reduces whenever a feature is used to split the data and gains importance then. The contribution will depend on all decision tress and can be normalized to give a relative importance score for each feature. Corresponding importance scores for each floor has been calculated to see which floors contributed most to damage classification and is given in Table 2. This showed that lower floors (especially Floors 1-3) had the most influential features with higher importance factor.

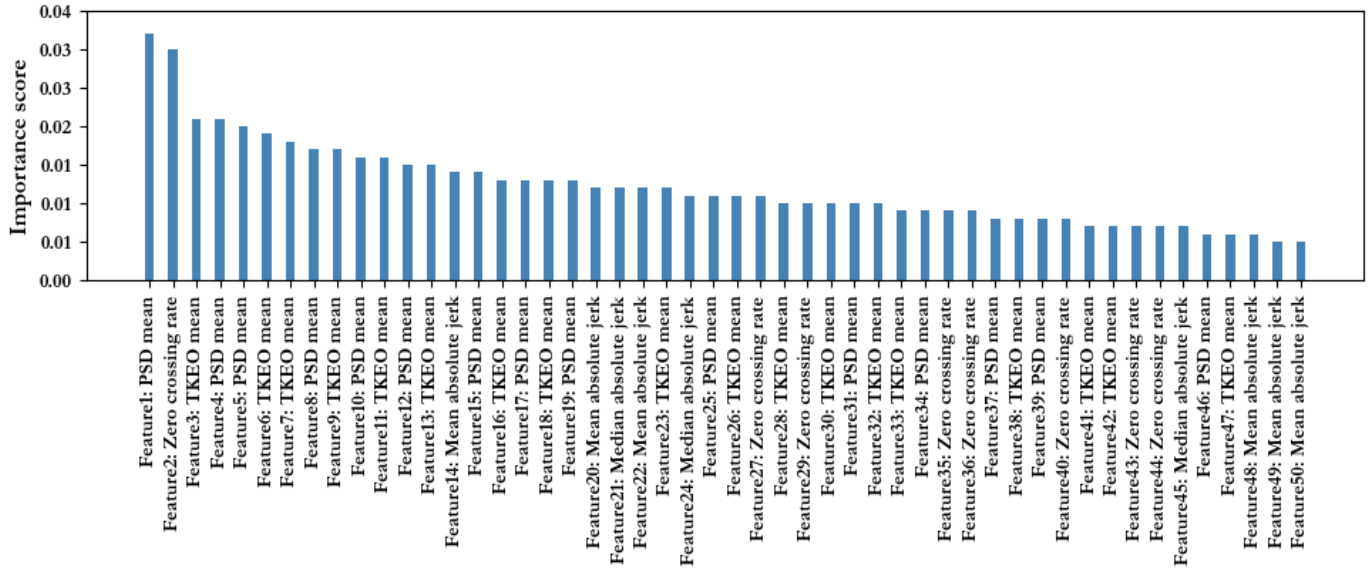


Figure 7

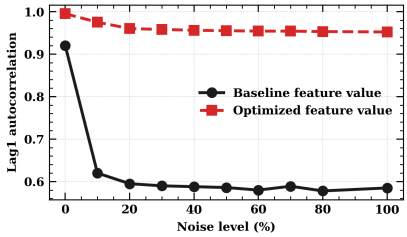
Table 2: Aggregated feature importance by floor

Floor	Total importance
Floor 1	0.088
Floor 2	0.086
Floor 3	0.081
Floor 4	0.074
Floor 5	0.070
Floor 6	0.066
Floor 7	0.059
Floor 8	0.057
Floor 9	0.053
Floor 10	0.051
Floor 11	0.048
Floor 12	0.047

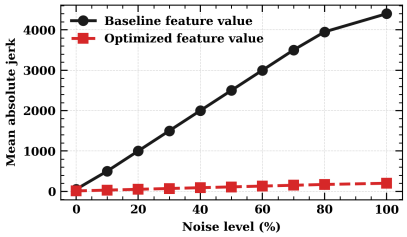
To quantitatively assess each floors significance in the damage classification task, an ablation-based sensitivity analysis was conducted. For that features corresponding to each floor was removed iteratively and the classification performance was further evaluated with overall accuracy metric.The mean decrease in impurity importance score for each floor and the corresponding accuracy drop showed a strong positive correlation coefficient ($\rho \approx 0.93$). Hence, the features, the model considered important was found to have measurable impact on prediction accuracy as well.

3.2 Effect of Noise on Optimized Features

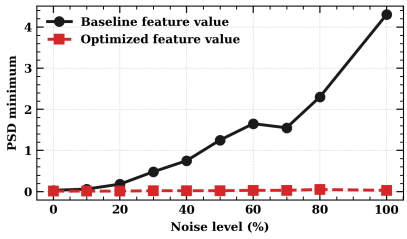
The effect of increasing levels of additive noise on each optimized feature was considered. For that, each feature is represented as a function $f(N)$, which shows how its average value changes with noise and $N \in [0, 100]\%$ represents the percentage of noise added to the time-series signal. The mean absolute jerk increased slowly and almost linearly with noise, characterized by a low gradient ($\frac{df}{dN} \approx b \ll 1$). The feature exhibited minimal sensitivity and consistent reliability even at high noise levels (up to $N = 100\%$).The zero crossing rate (ZCR) rapidly increased at lower noise levels but started to level off as noise became higher, following a curve with decreasing slope ($\frac{df}{dN} = Ake^{-kN}$). ZCR exhibited more sensitivity to low noise but remained stable up to around 70% noise. The lag-1 autocorrelation remained nearly invariant across all noise levels($\frac{df}{dN} \approx 0$). The power spectral density minimum (PSD_Min) was found to be stable at low noise levels but slight nonlinear increase suggesting some sensitivity beyond 60% noise was observed. These features are found to be minimally sensitive to noise which makes their suitability in real-world structural health monitoring (SHM) systems where sensor data is often contaminated with varying degrees of noise and are shown in Figure 8.



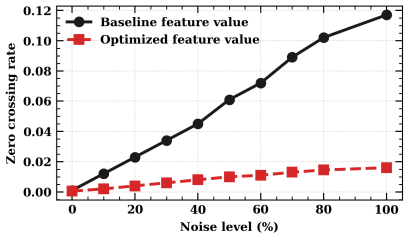
(a) Lag-1 autocorrelation



(b) Mean absolute jerk



(c) Minimum of the power spectral density



(d) Zero-crossing rate

Figure 8: Feature behavior as a function of noise factor for different signal characteristics: (a) Lag-1 autocorrelation, (b) Mean absolute jerk, (c) Minimum of the power spectral density , and (d) Zero-crossing rate.

3.3 Performance of Gradient Boosting Classifier

To understand the performance of gradient boosting classifier, the confusion matrix and misclassification rates (MCR) are considered and confusion matrix is shown in Figure 9. The model performs best for detecting low damage levels (especially at 10%), where the features are well-separated from other classes and data is less noisy. With increase in damage beyond 20%, features become harder to distinguish due to overlapping signal characteristics, leading to higher misclassification rates. Sharp decrease in model performance is quantified by the accuracy drop ratio (ADR), which measures the relative change in misclassification rate (MCR) between two consecutive damage classes and is given by,

$$\text{ADR}_{i \to i+1} = \left(\frac{\text{MCR}_i - \text{MCR}_{i+1}}{\text{MCR}_i} \right) \times 100 \tag{15}$$

where, MCR_i is the misclassification rate of class i, $\text{ADR}_{i \to i+1}$ shows the percentage increase or decrease in misclassification rate when moving from class i to $i + 1$. Accuracy drop ratio between consecutive classes for the building is given in Table 3. From the table, accuracy drop ratio states that the model performs best at lower damage levels and struggles most between 10% and 25%. This emphasizes the nonlinear impact of damage severity at moderate damage levels, where signal features become less distinct and affects classification accuracy.

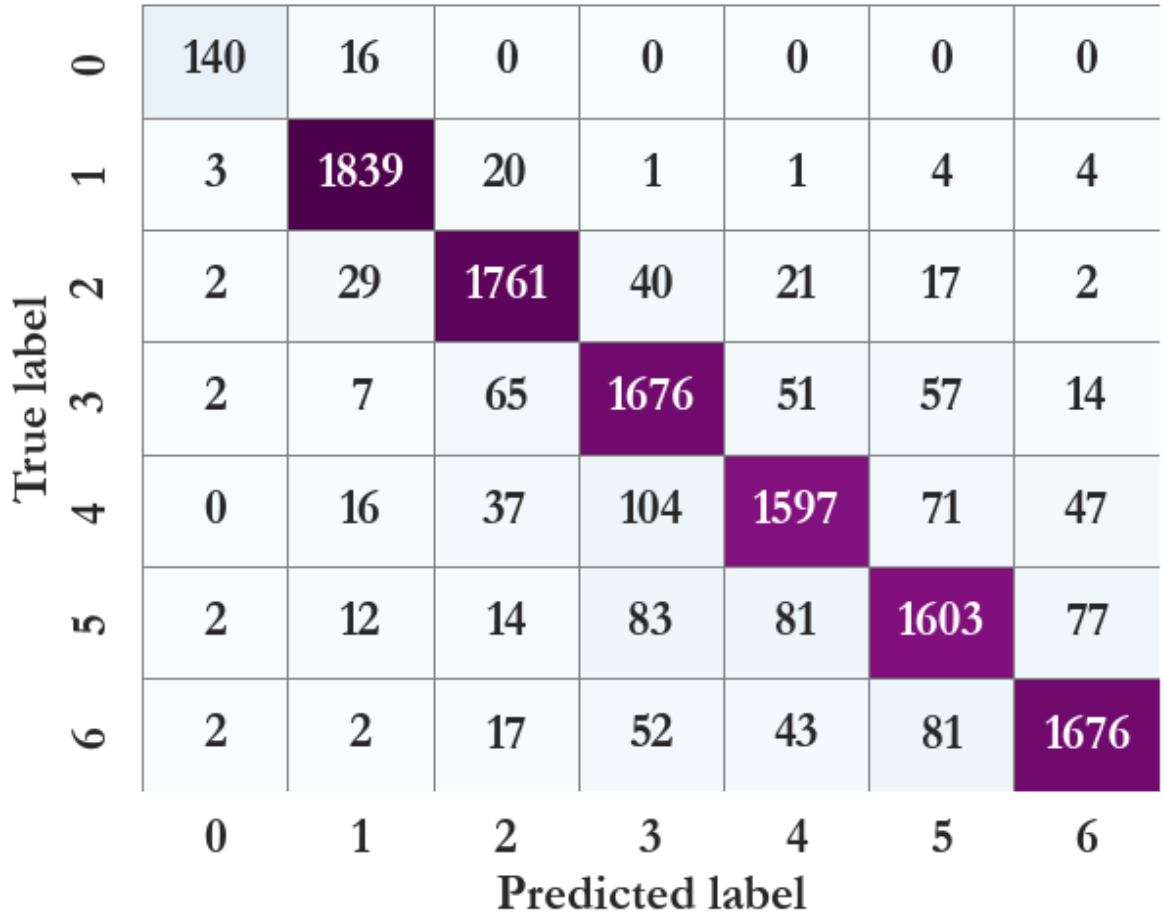


Figure 9: Confusion matrix of gradient boosting classifier

Table 3: Accuracy drop ratio (ADR) between consecutive classes

Transition	**MCR(i) (%)**	**MCR(i+1) (%)**	**ADR (%)**	**Interpretation**
$0 \rightarrow 1$ (5% $\rightarrow$ 10%)	10.26	1.71	-83.32	Improves
$1 \rightarrow 2$ (10% $\rightarrow$ 15%)	1.71	5.93	+246.78	Sharp decline in performance
$2 \rightarrow 3$ (15% $\rightarrow$ 20%)	5.93	10.47	+76.51	Moderate decline
$3 \rightarrow 4$ (20% $\rightarrow$ 25%)	10.47	14.69	+40.27	Continued decline
$4 \rightarrow 5$ (25% $\rightarrow$ 30%)	14.69	14.37	-2.18	Slight improvement
$5 \rightarrow 6$	14.37	10.47	-27.12	Noticeable improvement

3.4 Feature Importance Analysis

From Table 4, it is observed that most important features for predicting structural damage comes from both frequency-domain and time-domain signal analysis. Features derived from

power spectral density mean and zero-crossing rate contributed the most to the model's predictions. This indicates that both frequency-domain features (which captures energy distribution across different frequencies) and time-domain features (which reflects irregularities in signal patterns) play a crucial role in identifying structural damage levels.

Table 4: Aggregated feature importance by feature type

Feature type	**Total importance**
PSD mean	0.162
Zero crossing rate	0.147
TKEO mean	0.132
Lag1 autocorrelation	0.109
Median absolute jerk	0.101
Mean absolute jerk	0.093
STD mean	0.085
Median mean	0.071

3.4.1 Correlation Between Floor Importance and Classification Accuracy

To validate the relevance of the aggregated floor-wise importance values, an ablation-based sensitivity analysis was conducted. For each floor f, all associated features were removed from the model input, and the resulting drop in classification accuracy, denoted as ΔA_f, was recorded.

To quantify the correlation between the floor-wise importance I_f and accuracy degradation ΔA_f, Pearson correlation coefficient was computed and is given by,

$$\rho = \frac{\sum_{f=1}^{F}(I_f - \bar{I})(\Delta A_f - \overline{\Delta A})}{\sqrt{\sum_{f=1}^{F}(I_f - \bar{I})^2}\sqrt{\sum_{f=1}^{F}(\Delta A_f - \overline{\Delta A})^2}} \tag{16}$$

where $\bar{I}$ and $\overline{\Delta A}$ are the mean floor importance and mean accuracy drop, respectively, and F is the total number of floors.

A strong positive correlation, with $\rho \approx 0.93$, indicating that floors with higher feature importance also lead to greater drops in model performance when excluded, was observed. This result supports the reliability of the gradient boosting model' s internal attribution and reinforces that lower-floor features are critical to the model's predictive performance.

4 Results and Discussions

This study explored a noise-resilient approach to structural health monitoring (SHM) framework by integrating a carefully selected set of optimized signal features with machine learning to ensure reliable damage detection performance in challenging, real-world environments. The conclusions from the study includes:

1. A comprehensive range of multi-domain features, including statistical, spectral, autocorrelation, energy-based, and fractal metrics and rigorous evaluation of these for noise resilience are considered. The wavelet configurations that minimize feature variance across noise levels have been identified and the optimal noise invariance by considering the framework can enable stable performance during fault detection and predictive maintenance. Lag-1 autocorrelation was found to exhibit perfect robustness under high-noise environments. Mean absolute jerk and zero-crossing rate was found to be responsive to noise but maintained predictable and bounded behavior which is suitable for dynamic but controlled environments. Power spectral density minimum was found to be stable under moderate noise can be cautiously used in scenarios exceeding 60% noise contamination.

2. The analysis from 12-story shear building considering robust features shows that damage is best detected using data from the lower floors. A gradient boosting classifier is used alongside detailed feature-wise and floor-wise importance analyses, giving most critical contributors identifying floors 1 to 3 as key to accurate damage detection. Frequency and time based features (like spectral and zero-crossing) are found to be the most important based on importance factor. The gradient boosting classifier is found to demonstrate strong performance in identifying low damage levels, particularly at 10%, where features are found to be clearly distinguishable. But, as damage exceeds 20%, its accuracy declines noticeably due to increased feature overlap and reduced signal clarity. The gradient boosting model's feature importance scores accurately reflect real-world impact, confirming their value for guiding critical decisions like feature selection and sensor placement.

3. When existing structures are considered, safety and accountability are paramount. Considering a black-box approach which may offer strong predictive accuracy but not hinting

much into the physical phenomena is a serious issue with such approaches. In this study, gradient boosting model used here provides intrinsic feature importance scores, offering a quantitative understanding of which signal characteristics and sensor locations most influence the model's predictions. This approach makes it possible to align the algorithmic outcomes with engineering intuition which enable accurate damage detection along with data-driven decisions on sensor placement, feature selection, and system design optimization and makes the approach more interpretable.

This comprehensive approach highlights both accuracy and robustness of the method, making it more applicable for real-world SHM deployments. However, the performance of this adopted methodology in real operational environments needs to be validated since damage sensitive features are also sensitive to changes of environmental and operational conditions of the structure, which is beyond the scope of this study.

5 Data Availability Statement

The data, models, and code supporting the findings of this study are available from the corresponding author upon reasonable request.

6 Author Contributions

Ishan: Research planning, Data organization, Data interpretation, Method development, Code development, Results verification, Manuscript drafting. **Lakshmi Latha**: Manuscript drafting, Text editing and proofreading, Graphical presentation. **Samit Ray-Chaudhuri**: Idea development, Dataset organization, Research funding arrangement, Numerical work, Project coordination, Supervision, Results verification, Graphical presentation, Manuscript drafting, Text editing and proofreading.

7 Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sector.

8 Acknowledgements

References

[1] V.M. Karbhari and F. Ansari. *Structural Health Monitoring of Civil Infrastructure Systems*. Elsevier, New York, 2009.

[2] J. M. W. Brownjohn. Structural health monitoring of civil infrastructure. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1851):589–622, 2007.

[3] C.L. Wilson, K. Lonkar, S. Roy, F. Kopsaftopoulos, and F.-K. Chang. Structural health monitoring of composites. *7.20 Structural Health Monitoring of Composites*, pages 382–407, 2018.

[4] S.K. Thyagarajan, M.J. Schulz, and P.F. Pai. Detecting structural damage using frequency response functions. *Journal of Sound and Vibration*, 210(1):162–170, 1998.

[5] U. Lee and J. Shin. A frequency response function-based structural damage identification method. *Computers and Structures*, 80:117–132, 2002.

[6] J. He and Y. Zhou. A novel mode shape reconstruction method for damage diagnosis of cracked beam. *Mechanical Systems and Signal Processing*, 122:433–447, 2019.

[7] V. B. Dawari and G. R. Vesmawala. Structural damage identification using modal curvature differences. *IOSR Journal of Mechanical and Civil Engineering*, 4(43):33–38, 2013.

[8] B. Kotnik and Z. Kacic. A noise robust feature extraction algorithm using joint wavelet packet subband decomposition and ar modeling of speech signals. *Signal Processing*, 87 (6):1202–1223, 2007.

[9] O. Farooq and S. Datta. Wavelet-based denoising for robust feature extraction for speech recognition. *Electronics Letters*, 39(1):163–165, 2003.

[10] S. Hu, Y. Hu, X. Wu, J. Li, Z. Xi, and J. Zhao. Research of signal de-noising technique based on wavelet. *TELKOMNIKA Indonesian Journal of Electrical Engineering*, 11:5141–5149, 2013.

[11] A. Silik, M. Noori, Z. Wu, W. A. Altabey, J. Dang, and N. S. Farhan. Wavelet-based vibration denoising for structural health monitoring. *Urban Lifeline*, 2(1):1–14, 2024.

[12] Z. Hou, A. Hera, and M. Noori. Wavelet-based techniques for structural health monitoring. *Health Assessment of Engineered Structures: Bridges, Buildings and Other Infrastructures*, pages 179–202, 2013.

[13] M. Suchetha, N. Kumaravel, and B. Benisha. Denoising and arrhythmia classification using emd based features and neural network. In *2013 International Conference on Communication and Signal Processing*, pages 883–887. IEEE, 2013.

[14] C. Uyulan and T. T. Erguzel. Analysis of time-frequency eeg feature extraction methods for mental task classification. *International Journal of Computational Intelligence Systems*, 10(1):1280–1288, 2017.

[15] W. Caesarendra and T. Tjahjowidodo. A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing. *Machines*, 5(4):21, 2017.

[16] A.A. Jaber. Signal processing techniques for condition monitoring. *Design of an Intelligent Embedded System for Condition Monitoring of an Industrial Robot*, pages 53–73, 2017.

[17] O. Surucu, S. A. Gadsden, and J. Yawney. Condition monitoring using machine learning: A review of theory, applications, and recent advances. *Expert Systems with Applications*, 221:119738, 2023.

[18] X. Wei, D. Saha, and A. Quach. Exploiting multi-domain features for detection of unclassified electromagnetic signals. In *MILCOM 2024–2024 IEEE Military Communications Conference (MILCOM)*, pages 1–6. IEEE, 2024.

[19] D. Xu, H. Yang, C. Gu, Z. Chen, Q. Xuan, and X. Yang. Adversarial examples detection of radio signals based on multifeature fusion. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 68(12):3607–3611, 2021.

[20] D. Aiordachioaie. A comparative analysis of fault detection and process diagnosis methods based on a signal processing paradigm. *Discover Applied Sciences*, 7(1):10, 2024.

[21] B. F. Spencer Jr, S. H. Sim, R. E. Kim, and H. Yoon. Advances in artificial intelligence for structural health monitoring: A comprehensive review. *KSCE Journal of Civil Engineering*, 29(3):100203, 2025.

[22] T. G. Mondal and G. Chen. Artificial intelligence in civil infrastructure health monitoring-historical perspectives, current trends, and future visions. *Frontiers in Built Environment*, 8:1007886, 2022.

[23] F. G. Yuan, S. A. Zargar, Q. Chen, and S. Wang. Machine learning for structural health monitoring: challenges and opportunities. *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2020*, 11379:1137903, 2020.

[24] M. L. H. Souza, C. A. da Costa, G. de Oliveira Ramos, and R. da Rosa Righi. A feature identification method to explain anomalies in condition monitoring. *Computers in Industry*, 133:103528, 2021.

[25] Y. Lee, H. Kim, S. Min, and H. Yoon. Structural damage detection using deep learning and fe model updating techniques. *Scientific Reports*, 13(1):18694, 2023.

[26] R. E. Walpole and R. H. Myers. *Probability and Statistics for Engineers and Scientists*. Macmillan, New York, 1972.

[27] Pacific Earthquake Engineering Research Center (PEER). OpenSees - Open System for Earthquake Engineering Simulation. Available at: `http://opensees.berkeley.edu/`, PEER, Richmond, CA, USA.