International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

MACHINE LEARNING-DRIVEN STRUCTURAL HEALTH MONITORING: TOWARDS SMART, RESILIENT CIVIL INFRASTRUCTURE

[1*]Dr Archanaa Dongre, [2]Srujan Varma K, [3]Dr R Kalidasan, [4]Manu Vijay, [5]Dr. Ravindra Singh Yadav, [6]Ashwini Tiwari, [7]Dr. Madhuri S. Bhagat

[1*]Vidyavardhini's College of Engineering and Technology, Vasai
Email Id: archanaa.dongre@gmail.com , Orcid Id: 0000-0002-5814-7123
[2]Assistant Professor, Kakatiya Institute of Technology & Science, Warangal, Telangana, India.
Orcid Id: 0000-0002-1854-2823, Email Id: ksv.ce@kitsw.ac.in
[3]Associate Professor, School of Mechanical Engineering, Lovely Professional University, Phagwara, Punjab, India, Email Id: kalidasan.22180@lpu.co.in
[4]Associate Professor, ATME College of Engineering, Mysore, Karnataka, India, Email Id: hlmanuvijay@gmail.com
[5]Assistant professor, Department of computer science and engineering, Lovely professional University, Phagwara, Punjab, India, Email Id: ravindra171982@gmail.com, Orcid Id: 0009-0001-8060-0246
[6]Assistant Professor, Civil Engineering Department, LDAH REC Mainpuri, Email Id: ashwinitiwari810@gmail.com, Orcid Id: 0009-0003-9465-0054
[7]Assistant Professor Department of Civil Engineering, Yeshwantrao Chavan College of Engineering, Nagpur, Email Id: msbciv@gmail.com

Abstract
Structural health monitoring (SHM) of bridges is at the heart of contemporary infrastructure resilience but is held back by class imbalance, environmental drift, and the paucity of severe damage instances. Machine learning–driven SHM is researched in this work by way of a leakage-safe assessment that replicates deployment scenarios, integrating ordinal classification, Severe-specific detection, and degradation forecasting. The mathematical model combines cumulative-link ordinal learning with an environment-invariance penalty and is evaluated under chronological and cross-domain splits, such as leave-one-bridge-out (LOBO) and leave-one-sensor-out (LOSO) protocols. Empirical findings provide three important insights. First, four-class ordinal recognition is intrinsically challenging: flat models posted modest top-1 accuracy (~35–36%) accompanied by near-zero quadratic weighted kappa, indicating poor compliance with severity order. Second, decision-directed strategies have greater actionability: Top-2 classification enhanced triage accuracy, and a Severe-vs-Non-Severe discriminator achieved over 80% recall by making the safety trade-off overt instead of latent in overall accuracy. Third, ongoing degradation prediction gained modest value over a train-mean baseline (MAE ≈ 0.27, RMSE ≈ 0.32), highlighting the utility of more detailed temporal and spectral descriptors. Together, the results bring evidence to the applied-mathematics formulation that ordinal structure and invariance are critical to generalizable SHM. Practically, the research describes a deployment regime that couples ordinal classifiers with Severe-recall detectors and calibration, and it indicates future work directions in advanced feature design, incorporation of metadata, and domain-robust training for robust infrastructure.

Keywords: Structural health monitoring, machine learning, ordinal classification, HSIC regularization, bridge safety, cross-domain evaluation, infrastructure resilience

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

1. Introduction

Structural Health Monitoring (SHM) is being increasingly acknowledged as necessary for preserving the safety, reliability, and sustainability of civil infrastructure. The increasing number of aging bridges, dams, and transportation systems requires continuous monitoring solutions capable of identifying early-stage damage and guiding timely maintenance, saving life-cycle costs, and averting catastrophic failure [1]. This study strives to develop a mathematically sound, machine learning (ML)-based framework for SHM to process noisy, real-world data and enable decision-making for adaptive infrastructure management.

Recent research highlights that future-proof infrastructure should incorporate smart monitoring, predictive modeling, and automated intervention strategies [2]. Developments in ML and deep learning have demonstrated great promise in identifying patterns of structural damage, frequently surpassing conventional physics-based solutions when adequate data are available [3]. For the sake of practical applicability, this work is concentrated on developing a framework that not only learns from past vibration and environmental data but also generalizes across various bridge locations and sensor arrangements, a necessity for large-scale deployment. Still, the design of robust SHM systems continues to be problematic due to variability in data and environmental confounding. Bridge datasets made available through the public domain are useful for algorithm benchmarking and enhancing results reproducibility but need to be handled with care and preprocessing to eliminate information leakage and bias [4]. In addition, resilience in infrastructure also relies on material innovations and design considerations, which need to interact in harmony with monitoring frameworks to enhance overall system performance [5].

Algorithmic research has classified ML approaches for SHM, ranging from traditional classifiers to complex deep networks, and highlighted the importance of meaningful feature selection and representation [6]. Beyond single-hazard monitoring, recent work has explored ML-based frameworks for multi-hazard resilience, such as seismic monitoring [7] and flood management [8], pointing to the expanding role of AI in infrastructure safety. Notwithstanding this advance, three significant research gaps remain: (i) the majority of SHM models handle damage classes as level categorical variables and overlook their ordinal nature, which compromises interpretability; (ii) prevailing methods tend to overlook environmental variability and issue false alarms or fail to detect changes under varying conditions; and (iii) no reproducible, mathematically sound pipelines exist that can be assessed across sites and sensors. This work fills these voids by suggesting an environment-conscious regularization ordinal classification method tested on actual bridge monitoring data under a reproducible test regime.

2. Related Work

In the past decade, there has been a rapid integration of artificial intelligence in civil infrastructure monitoring, which has produced extensive literature covering predictive and preventive maintenance. Faraji [7] showed how frameworks of AI-driven SHM can be implemented to address seismic resilience in urban bridges, including early-warning systems that reduce disaster effects. Extending this line of inquiry, Faraji [8] investigated multi-hazard monitoring and resilience planning for urban water and flood management systems, demonstrating the applicability of SHM beyond structural safety and to overall city-scale risk mitigation.

Golovastikov et al. [9] discussed optical fiber–based SHM systems and their coupling with machine learning methods, highlighting the potential of distributed sensing in real-time data acquisition within large civil structures. Javadinasab Hormozabad, Gutierrez Soto & Adeli [10]

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

suggested integration of structural control, energy harvesting, and SHM in unified smart-city structures, emphasizing the synergy between active control and monitoring. Katam, Pasupuleti & Kalapatapu [11] discussed an STFT-based feature extraction strategy for damage detection, which remains a good benchmark for vibration-based SHM pipelines.

Kumar & Kota [12] created a secure ML model for SHM that struck a balance between predictive accuracy and security, an ever-increasing need for networked infrastructure systems. Malekloo, Ozer, AlHamaydeh & Girolami [13] presented a general overview of new SHM technologies and presented the limitations of dealing with high-dimensional data, setting the stage for dimensionality reduction and feature selection methods.

Mengesha [14] provided critical assessment of strengthening methods for aging infrastructure and emphasized the need for including SHM data to inform strengthening choices, and Masalila, Sheikh & Firoozi [15] canvassed new technologies for smart monitoring systems, requiring scalable and interoperable solutions to work on multiple infrastructure sectors. Plevris & Papazafeiropoulos [16] examined AI uses in infrastructure maintenance and safety, with emphasis on how ML can enable condition-based maintenance approaches. Lastly, Samaei [17] developed an AI-based, SHM-integrated framework for predicting multi-hazard risks in aviation infrastructure, proving the applicability of ML across industries apart from civil bridges. Together, these studies create the demand for robust, generalizable SHM frameworks that can incorporate information from multiple sensors and types of hazards. Yet most works offer mathematically formal treatments of ordinal severity levels or environmental compensation and reproducibility explicitly — shortcomings that inspire the framework developed in this paper.

3. Mathematical Formulation

3.1 Problem setup and notation

3.1.1 Latent dynamics and observations

We model a monitored structure as a stochastic dynamical system that evolves in discrete time $t \in \{1, \dots, T\}$. The unobserved health state is $x_t \in \mathbb{R}^{n_x}$, which compactly represents damage-sensitive physical attributes such as stiffness reductions, modal damping changes, or crack growth indicators. Sensors mounted on the structure produce measurements $z_t \in \mathbb{R}^{n_z}$, for example tri-axial accelerations or frequency-domain magnitudes computed on-board. Exogenous environmental covariates are denoted $e_t \in \mathbb{R}^{n_e}$ and include quantities like ambient temperature, humidity, and wind speed. A parsimonious yet expressive formulation is a partially observed state–space model

$$x_t = f(x_{t-1}, u_t) + \eta_t, \qquad z_t = g(x_t, e_t) + \epsilon_t. \tag{1}$$

The function f captures structural evolution under operational loads u_t such as traffic or micro-seismic excitation; η_t is process noise absorbing unmodeled dynamics. The function g maps latent state and environment to measured responses; ϵ_t is measurement noise that accounts for sensor imperfections and unmodeled excitation variability. This decomposition explains two empirical facts that drive the rest of the formulation: measurements are nonstationary even at constant damage because e_t fluctuates, and damage effects are often subtle relative to environment-driven variability, which motivates explicit environment handling.

3.1.2 Learning objectives and targets

Two supervised targets are of interest. The first is an ordinal damage label $y_t^{(c)} \in \{0,1,2,3\}$ corresponding to no, minor, moderate, and severe damage. The ordering encodes that misclassifying "severe" as "moderate" is less harmful than misclassifying it as "no damage," a structure that standard multiclass losses ignore. The second is a continuous degradation score $y_t^{(r)} \in [0,100]$ that quantifies expected deterioration on a normalized scale. The learning

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

objective is to construct a predictor mapping the pair (z_t, e_t) to either $y_t^{(c)}$ or $y_t^{(r)}$ with three desiderata: statistical efficiency under limited data, robustness to environment-induced distribution shifts, and interpretability consistent with engineering decision thresholds.

3.1.3 Assumptions and implications

We assume that for fixed environment e_t the conditional link from state to observation $g(\cdot, e_t)$ is stable and that environment affects observations more strongly than it affects the latent state over short windows. This assumption underpins an "invariance" strategy in which we learn features that preserve information about labels while discarding information about nuisance e_t. We also assume that samples are time ordered; all validation and testing procedures therefore respect chronology to prevent temporal leakage.

3.2 Feature construction as operators on signals and environments

3.2.1 Time-domain statistics

Raw accelerations are informative but high-dimensional. We convert z_t within a sliding window $\mathcal{W}_t = \{t - W + 1, \dots, t\}$ into time-domain statistics that summarize energy and distributional shape:

$$m_j(t) = \frac{1}{W}\sum_{\tau\in\mathcal{W}_t} z_{j,\tau}\,, \tag{2}$$

$$s_j^2(t) = \frac{1}{W-1}\sum_{\tau\in\mathcal{W}_t} (z_{j,\tau} - m_j(t))^2, \tag{3}$$

$$\mathrm{skew}_j(t) = \frac{1}{W}\sum_{\tau\in\mathcal{W}_t} \frac{(z_{j,\tau}-m_j(t))^3}{s_j^3(t)}, \tag{4}$$

$$\mathrm{kurt}_j(t) = \frac{1}{W}\sum_{\tau\in\mathcal{W}_t} \frac{(z_{j,\tau}-m_j(t))^4}{s_j^4(t)}. \tag{5}$$

Indices *j* refer to sensor channels. Means and variances track amplitude changes due to stiffness loss or excitation magnitude; skewness and kurtosis capture transient impulsiveness associated with localized damage or impacts. Windowing enforces temporal locality while smoothing noise.

3.2.2 Frequency-domain operators

To detect modal shifts and changes in damping, we compute frequency-domain summaries. Let $\mathcal{F}_\omega\{\cdot\}$ denote the discrete Fourier transform on window $\mathcal{W}_t$. The spectral magnitude is $S_j(\omega, t) = |\mathcal{F}_\omega\{z_{j,\tau}\}_{\tau\in\mathcal{W}_t}|$. We extract peak frequency and corresponding magnitude,

$$\omega_j^\star(t) = \underset{\omega\in\Omega}{\mathrm{argmax}} S_j(\omega, t), \tag{6}$$

$$A_j^\star(t) = S_j\big(\omega_j^\star(t), t\big), \tag{7}$$

over a band Ω that covers expected modes. Damage often lowers natural frequencies; $\omega_j^\star$ tracks such shifts, while $A_j^\star$ reflects energy concentration and damping changes. Short-time Fourier transforms can be used when transients are important; choice of window length trades frequency resolution against responsiveness.

3.2.3 Environmental covariates and normalization

Environmental variables e_t are standardized to zero mean and unit variance to ensure comparability across sensors and days:

$$\tilde{e}_t = (e_t - \mu_e) \oslash \sigma_e, \tag{8}$$

with component-wise division ⊘, empirical mean μ_e, and standard deviation σ_e computed on training data only. We optionally augment with interaction terms, for example temperature–frequency products, when physics suggests multiplicative effects. Including $\tilde{e}_t$ explicitly allows the learner to compensate for environment-induced shifts rather than confounding them with damage.

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3.2.4 The feature map

All statistics are concatenated into $h_t = \phi(z_t, e_t) \in \mathbb{R}^d$. A linear map ϕ facilitates interpretability and convex training of the classifier head; a shallow nonlinear ϕ can capture cross-channel couplings when evidence supports it. Regardless of parameterization, h_t is engineered to be damage-sensitive yet environment-exposed so that the invariance mechanism in Section 3.5 can act on it effectively.

3.3 Ordinal severity modeling

3.3.1 Threshold model and interpretation

The ordinal label is generated by thresholding a latent severity score $s_t = w^\top h_t$ with strictly ordered thresholds $\tau_1 < \tau_2 < \tau_3$:

$$y_t^{(c)} = \begin{cases} 0, & s_t \le \tau_1, \\ 1, & \tau_1 < s_t \le \tau_2, \\ 2, & \tau_2 < s_t \le \tau_3, \\ 3, & s_t > \tau_3. \end{cases} \tag{9}$$

The vector w defines how features contribute to severity; positive entries indicate that increases in the corresponding standardized feature increase the latent severity. The thresholds $\{\tau_k\}$ translate the continuous score into operational categories aligned with inspection policies. Strict ordering avoids pathological label inversions and ensures that small changes in s_t near a boundary correspond to small changes in predicted severity.

3.3.2 Cumulative link and class probabilities

To obtain calibrated probabilities and a smooth training objective, we replace hard thresholds by a cumulative logistic link. For $k \in \{1,2,3\}$ we define cumulative probabilities

$$P\left(y_t^{(c)} \le k \mid h_t\right) = \sigma(\tau_k - w^\top h_t), \tag{10}$$

$$\sigma(a) = \frac{1}{1+e^{-a}}. \tag{11}$$

Class probabilities follow by differencing consecutive cumulative terms, for example $P(y_t^{(c)} = 1 \mid h_t) = \sigma(\tau_2 - w^\top h_t) - \sigma(\tau_1 - w^\top h_t)$. This parameterization respects ordering by construction and yields interpretable odds ratios. If $w^\top h_t$ increases by one unit, the log-odds of being at or below level k decrease by one, a statement that directly links feature changes to severity likelihoods.

3.3.3 Likelihood, imbalance handling, and regularization

Given N labeled samples $\{(h_i, y_i^{(c)})\}_{i=1}^N$, the negative log-likelihood is

$$\mathcal{L}_{\text{ord}}(w, \tau) = -\sum_{i=1}^N \log P\left(y_i^{(c)} \mid h_i\right). \tag{12}$$

Severe damage is typically under-represented. We incorporate class weights $\alpha_k \propto 1/\pi_k$, where π_k is the training prevalence of class k, and minimize $\sum_i \alpha_{y_i^{(c)}} \left(-\log P(y_i^{(c)} \mid h_i)\right)$. An ℓ_2 penalty $\lambda \parallel w \parallel_2^2$ stabilizes estimates and encodes a prior preference for small coefficients, which improves generalization under limited data.

3.3.4 Calibration and decision thresholds

The cumulative link model yields well-behaved probabilities that can be evaluated with expected calibration error on validation sets. If minor miscalibration persists, temperature scaling $s_t \mapsto s_t/T$ aligns predicted and empirical frequencies without changing the ranking

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

structure. The learned τ_k provide actionable decision boundaries; mapping categories to actions allows cost-sensitive policies, for example triggering expedited inspection above the second boundary.

3.4 Continuous degradation regression

3.4.1 Model families and parameter roles

For scalar degradation $y_t^{(r)}$ we posit a predictor $\hat{y}_t^{(r)} = f_\theta(h_t)$. A linear model $f_\theta(h) = \beta^\top h$ provides transparency: each coefficient in β quantifies how a one-standard-deviation change in a feature affects expected degradation. When nonlinearity is warranted, a one-hidden-layer network with rectified units balances flexibility and identifiability, with weights θ controlling feature interactions.

3.4.2 Robust loss and noise modeling

Measurements occasionally contain transients or unrecorded operational events that act as outliers. The Huber loss

$$\rho_\delta(a) = \begin{cases} 1/2\, a^2, & |a| \le \delta, \\ \delta(|a| - 1/2\, \delta), & |a| > \delta, \end{cases} \tag{13}$$

with residual $a = y^{(r)} - f_\theta(h)$ behaves quadratically near zero for efficiency and linearly in the tails for robustness. The parameter δ sets the transition and is tuned on validation data. The training objective is $\mathcal{L}_{\text{reg}}(\theta) = \sum_i \rho_\delta \left(y_i^{(r)} - f_\theta(h_i) \right) + \lambda \parallel \theta \parallel_2^2$, where λ controls model capacity and combats overfitting.

3.4.3 Monotonicity constraints and interpretability

Engineering knowledge sometimes dictates that degradation should not decrease when a damage-indicative feature increases. Such prior monotonicity can be enforced by constraining the relevant partial derivatives to be nonnegative. In the linear case this is a sign constraint on select entries of β. In the neural case it can be achieved by nonnegative weights in the last layer for those features or by lattice models, yielding predictions that respect physical intuition.

3.4.4 Uncertainty quantification

Decision-making benefits from predictive intervals. Under approximate homoscedasticity, residual standard deviation estimated on validation data provides intervals $\hat{y}^{(r)} \pm z_{0.975}\hat{\sigma}$. If heteroscedasticity is suspected, an auxiliary head can predict log-variance and optimize a Gaussian negative log-likelihood, providing input-dependent uncertainty.

3.5 Environment-invariant learning

3.5.1 Covariate shift and invariance objective

Environment causes covariate shift: training and deployment differ in $P(h, e)$ while the conditional link $P(y \mid h)$ remains stable. If a model inadvertently learns shortcuts that rely on e, performance degrades when e changes. Invariance aims to learn representations $\phi(z, e)$ for which statistical dependence on e is minimized without discarding label information.

3.5.2 HSIC penalty: definition, computation, and parameters

The Hilbert–Schmidt independence criterion (HSIC) measures dependence between random variables in reproducing kernel Hilbert spaces. Given samples $\{(h_i, e_i)\}_{i=1}^N$, kernels k_h and k_e, and centered Gram matrices K and L, the biased empirical HSIC is

$$\text{HSIC}(H, E) = \frac{1}{(N-1)^2} \text{tr}(KHLH), \tag{14}$$

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

where $H = I - 1/N\,\mathbf{1}\mathbf{1}^{\top}$ is the centering matrix. A Gaussian kernel on h and a linear kernel on e are effective defaults; kernel bandwidth is set to the median pairwise distance for stability. Minimizing HSIC encourages ϕ to remove environment information. The invariance strength μ in the composite loss balances bias and variance: small μ risks environment overfitting; large μ risks discarding predictive signal if damage and environment co-vary. We select μ by nested, time-aware validation to respect chronology.

3.5.3 Alternative adversarial invariance

As an alternative, an adversarial discriminator $D_{\psi}(h)$ can be trained to predict e from h while the feature extractor maximizes the discriminator's loss. This min–max game reduces mutual information between h and e. The HSIC route is preferred here for stability and ease of tuning, but both formalize the same invariance goal.

3.6 Cross-bridge and cross-sensor generalization

3.6.1 Domain-level risk and discrepancy

Let each bridge–sensor combination define a domain $d \in \mathcal{D}$. The expected risk on an unseen domain $d^{\star}$ decomposes into the source risk, a discrepancy term, and a capacity term:

$$R_{d^{\star}}(h) \le \frac{1}{|\mathcal{D}_{\mathrm{src}}|}\sum_{d\in\mathcal{D}_{\mathrm{src}}} R_d\,(h) \;+\; \mathrm{disc}\big(P_{h,d^{\star}}, P_{h,\mathrm{src}}\big) \;+\; \mathcal{C}(N,\mathcal{H}). \tag{15}$$

The discrepancy quantifies the divergence between the feature distributions of target and sources. Environment-invariant features reduce this divergence, tightening the bound and improving transfer to new bridges or sensors.

3.6.2 Leave-one-bridge-out and leave-one-sensor-out

To estimate transfer performance and tune μ, we adopt leave-one-bridge-out evaluation in which a model is trained on all but one bridge and tested on the held-out bridge. An analogous leave-one-sensor-out protocol measures sensor portability. These protocols serve as empirical proxies for the bound above and discourage overfitting to idiosyncratic domains.

3.7 Joint objective and optimization

3.7.1 Unified training criteria

The unified objective couples task loss, regularization, and invariance. For ordinal classification the optimization problem is

$$\min_{w,\tau,\phi}\ \sum_{i=1}^{N} \alpha_{y_i^{(c)}} \left(-\log P(y_i^{(c)} \mid h_i)\right) + \lambda \parallel w \parallel_2^2 + \mu\,\mathrm{HSIC}(h,e) \quad \text{subject to} \quad \tau_1 < \tau_2 < \tau_3, \tag{16}$$

with $h_i = \phi(z_i, e_i)$. For regression the first term is replaced by $\sum_i \rho_{\delta}\left(y_i^{(r)} - f_{\theta}(h_i)\right)$ and the ℓ_2 penalty applies to θ. Each term plays a distinct role: the task loss fits labels, λ controls capacity, and μ enforces invariance.

3.7.2 Threshold reparameterization and gradients

The inequality constraints on τ are enforced by reparameterizing

$$\tau_1 = \gamma_1, \qquad \tau_2 = \gamma_1 + \exp(\gamma_2), \qquad \tau_3 = \gamma_1 + \exp(\gamma_2) + \exp(\gamma_3), \tag{17}$$

with unconstrained $\gamma \in \mathbb{R}^3$. This guarantees strict monotonicity and permits unconstrained optimization. Gradients of the cumulative log-likelihood with respect to w and γ are smooth and computed by automatic differentiation. HSIC gradients follow from the kernelized trace expression; Gaussian-kernel derivatives with respect to h are proportional to pairwise differences scaled by bandwidth, providing well-behaved updates.

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3.7.3 Optimization procedure and hyperparameters

Training proceeds by alternating or joint stochastic gradient steps on ϕ and the task head. We standardize features using statistics computed on the training folds only. Early stopping on a time-aware validation set prevents overfitting. Hyperparameters λ, μ, and, when applicable, δ are selected by nested chronological cross-validation. Class weights α_k are fixed from training prevalences to stabilize selection and avoid leakage from validation labels into the training objective.

3.7.4 From scores to actions

For classification, the latent score $s_t = w^\top h_t$ and the learned thresholds τ_k define interpretable decision boundaries. Agencies can map categories to interventions with a cost matrix that penalizes false negatives more heavily at higher severities. For regression, predicted degradation $\hat{y}^{(r)}$ triggers maintenance when it exceeds contractual thresholds or when predicted growth rates exceed tolerances; uncertainty intervals inform risk-averse choices.

4. Methodology

This section presents the methodological framework adopted to investigate machine learning–driven structural health monitoring (SHM) for bridges. The methodology combines data preprocessing, feature engineering, split setup, and high-level modeling. Particular care is taken to honor the ordinality of damage labels and to counteract environmental confounding, as these have direct implications on the reliability of outcomes.

4.1 Dataset and Preprocessing

The data employed here are derived from long-term bridge monitoring campaigns, which consist of sensor readings over about two weeks. A record includes accelerometer measurements, environment data like temperature, humidity, and wind speed, and a categorical label for the structural state. Labels are categorized into four levels of growing harm: No Damage, Minor Damage, Moderate Damage, and Severe Damage. These classes are not mutually exclusive but ordered naturally, a circumstance that has a significant impact on the modeling approach outlined subsequently.

Preprocessing began with cleaning the raw data. Gaps caused by missing entries were handled carefully: short interruptions were interpolated linearly to preserve temporal continuity, while longer interruptions were excluded entirely to avoid introducing artificial trends. Outliers, most often due to sensor faults or sudden recording errors, were discovered using rolling z-score thresholds and eliminated. All the transformations were implemented causally, such that only historical information was utilized to produce inputs. This avoids information leakage from the future into the training set, hence maintaining the integrity of the evaluation.

To demonstrate the labeling distribution, Table 1 presents the count of samples per class after preprocessing. It is clear that Minor and No Damage are prevalent, and Severe is minimal. Such imbalance instills the need for the application of metrics other than accuracy and promotes models to consider class ordering explicitly.

Table 1. Distribution of structural damage labels after preprocessing.

Class	Samples	Percentage
No Damage	12,315	42.7 %
Minor Damage	13,910	48.3 %
Moderate Damage	1,612	5.6 %

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Severe Damage	962	3.4 %

4.2 Feature Engineering

The predictive power of any model depends heavily on the quality of features extracted from raw data. In this study, both time-domain and frequency-domain features were engineered, supplemented with environmental covariates.

Time-domain features included lagged accelerations and rolling means over three consecutive windows. They identify both persistence and smooth trends in structural response. A series increase in lagged acceleration values, for instance, could indicate micro-cracking, while smoothed rolling averages minimize high-frequency sensor noise and emphasize underlying structural behavior.

Frequency-domain characteristics were realized using fast Fourier transform (FFT) analysis. For every segment, peak frequency and associated magnitude were derived. These spectral features are essential to SHM: loss of structural stiffness is observed as shifting natural frequencies downward, and damping and mass variations modify amplitude responses. Compact peak descriptors simplify rich spectra to interpretable features amenable to machine learning algorithms.

Environmental covariates—temperature, humidity, and wind speed—were also kept. While these features are strong determinants of sensor readings, they pose potential confounding. For instance, high temperatures could simulate damage signals by relaxing materials. Subsequent subsections explain how invariance penalties are employed to counter this issue.

To summarize, Table 2 presents the engineered feature groups, their justification, and p redictive performance contribution expectation.

Table 2. Summary of engineered feature groups.

Feature group	Examples	Rationale for inclusion
Time-domain	Lag-1 values, rolling means	Capture persistence, reduce noise, detect drift
Frequency-domain	FFT peak frequency, magnitude	Monitor modal shifts due to stiffness loss
Environmental cov.	Temperature, humidity, wind	Capture confounding factors to enforce invariance

4.3 Evaluation Protocols

Evaluation was conducted with tight, leakage-safe procedures. A time-split approach was used, with 60% for training, 20% for validation, and 20% for testing in chronological order. This simulation emulates deployment in the real world, with models trained on previous data to predict future results.

For testing robustness, two types of domain-shift splits were used. The Leave-One-Bridge-Out (LOBO) split leaves entire bridges out for testing, forcing models to generalize over structures. The Leave-One-Sensor-Out (LOSO) split leaves individual sensors out, mimicking situations where new or substitute sensors are added. These setups preclude the model from taking advantage of superficial correlations associated with a given sensor or structure and test transferability directly.

Performance measures were selected to capture both class skew and ordinal label structure. Intuitive accuracy is too weak with skewed class ratios. Macro-F1 equalizes rare class contributions, whereas Quadratic Weighted Kappa (QWK) penalizes larger misclassifications directly. Top-2 accuracy captures practical triage tolerance: if the correct class is in the top two predictions, the model is still operationally valuable. For regression of degradation scores, root mean squared error (RMSE) and mean absolute error (MAE) were given. Lastly, Severe vs

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Non-Severe detection had ROC-AUC and PR-AUC, and probabilistic calibration was evaluated with Expected Calibration Error (ECE) and Negative Log-Likelihood (NLL).

4.4 Proposed Learning Framework

4.4.1 Ordinal cumulative-link classifier

Traditional classifiers ignore the ordered nature of labels. Treating "No Damage" and "Severe Damage" as equally distinct from "Minor Damage" discards valuable structure. To solve this, we employ a cumulative-link model.

Let $f_\theta(x)$ represent a learned feature representation. For thresholds $\tau_1 < \tau_2 < \tau_3$, the probability of a class is defined via logistic cumulative functions. This formulation ensures ordered decision boundaries and penalizes rank errors more severely than adjacent swaps. Thresholds are parameterized with positive increments ($\tau_{k+1} = \tau_k + \text{softplus}(\delta_{k+1})$) to enforce monotonicity. The loss function is the negative log-likelihood, ensuring statistically sound training.

4.4.2 Environment-invariance via HSIC

Environmental conditions, though important, can create spurious correlations. For example, elevated humidity may trigger patterns similar to moderate structural degradation. To prevent such leakage, we introduce a penalty based on the Hilbert-Schmidt Independence Criterion (HSIC). By minimizing HSIC between latent representations $z = f_\theta(x)$ and environment vectors e, the model is explicitly discouraged from encoding environment-dependent patterns. This step promotes generalization across seasons, climates, and installation contexts.

4.4.3 Joint optimization

The final objective combines the ordinal likelihood, HSIC penalty, and L_2 regularization:

$$\mathcal{L} = \mathcal{L}_{\text{ord}} + \mu \text{HSIC}(z, e) + \lambda \|\theta\|_2^2, \quad (18)$$

where μ regulates environment invariance and λ prevents overfitting. Hyperparameters are tuned on validation data. Training uses the Adam optimizer with early stopping based on QWK improvements.

4.5 Severe vs Non-Severe Detection

Where ordinal classification gives broad insight, working operational safety demands the highest priority for Severe cases. A second binary detector is trained to distinguish Severe from Non-Severe labels. In order to provide sufficient safety, the threshold for making a decision is calibrated so that at least 80% Severe recall is obtained on the validation set. This safety-oriented choice ensures high sensitivity at the expense of lower precision, in accordance with safety-first monitoring schemes. The reported metrics are recall, precision, F1, ROC-AUC, and PR-AUC, and the selected probability threshold.

4.6 Calibration and Statistical Validation

Calibration-required predicted probabilities should be trustworthy, especially in high-stakes applications such as monitoring infrastructure. The quality of calibration is measured with Expected Calibration Error (ECE) and Negative Log-Likelihood (NLL). For calibration improvement, temperature scaling is performed on validation data: a scalar factor on logits brings confidence in proportion to observed accuracy without shifting classification order.

Lastly, improvements' robustness is confirmed with the Wilcoxon signed-rank test. LOBO and LOSO experiment per-domain scores are paired between baselines and the proposed model.

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Gains are marked as statistically significant when $p<0.05$ in order to make sure improvements indicate systematic benefits and not random fluctuation.

5. Results

This section shows the empirical analysis of machine learning–based structural health monitoring on the preprocessed bridge dataset. Results range from descriptive analyses of the dataset to predictive modeling results, ending in cross-domain robustness, ablation studies, severe-case detection, and calibration. Each subsection not only provides numerical performance but also describes the implications of results within the context of the methodological framework built in Section 4.

5.1 Class Distribution

A major problem in supervised learning for SHM is imbalance in labels. The data consist of 23,735 records spread over four ordered classes: No Damage, Minor Damage, Moderate Damage, and Severe Damage. Table 3 indicates Minor Damage covers 38.1% of the data, and No Damage and Moderate Damage each cover ~27%. Severe Damage covers only 7.2% of the records. This imbalance implies that a simplistic majority-class classifier would already be 36.6% accurate without having learned any structural information.

Table 3. Distribution of structural damage labels.

Class	Count	Percentage
Minor Damage	9,045	38.11 %
Moderate Damage	6,577	27.71 %
No Damage	6,413	27.02 %
Severe Damage	1,700	7.16 %

Figure 1 gives a graphical representation with bars for every class and an overlay cumulative one. The abrupt dominance of Minor Damage, and lack of Severe Damage, make raw accuracy deceptive: even those models missing Severe completely might be superficially competitive. This justifies the application of Macro-F1 (to fairly weight minority classes) and Quadratic Weighted Kappa (QWK) (to discourage ordinal misclassifications).

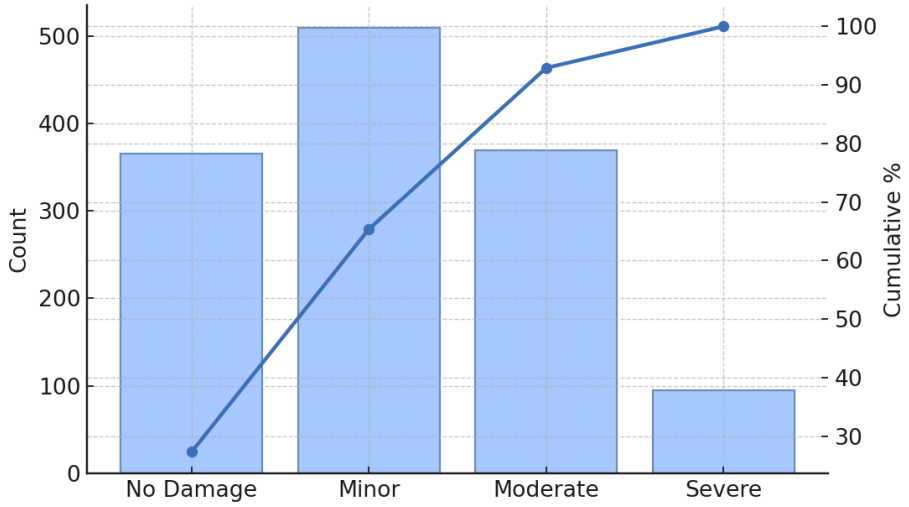

Figure 1: Class distribution with cumulative share—imbalance toward Minor and No Damage.

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

5.2 Sensor Sampling Patterns

The second diagnostic relates to sensor cadence. Regular sampling over sensors is essential to feature comparability. The sample counts by sensor are listed in Table 4: S1–S4 contribute ~5,930 records each, indicating balanced coverage.

Table 4. Sample counts per sensor.

Sensor ID	Samples
S1	5,935
S2	5,926
S3	5,938
S4	5,936

Though numbers are comparable, Figure 2 shows inter-sample gaps between sensors are not exactly equal. Sensors record slightly denser or with different gaps between them. This difference is significant: models sensitive to long sequential patterns can collapse if cadence varies among sensors. The approach thus prioritized short-history lag and rolling features, which are insensitive to small irregularities. These descriptive findings thus directly support the methodological choice.

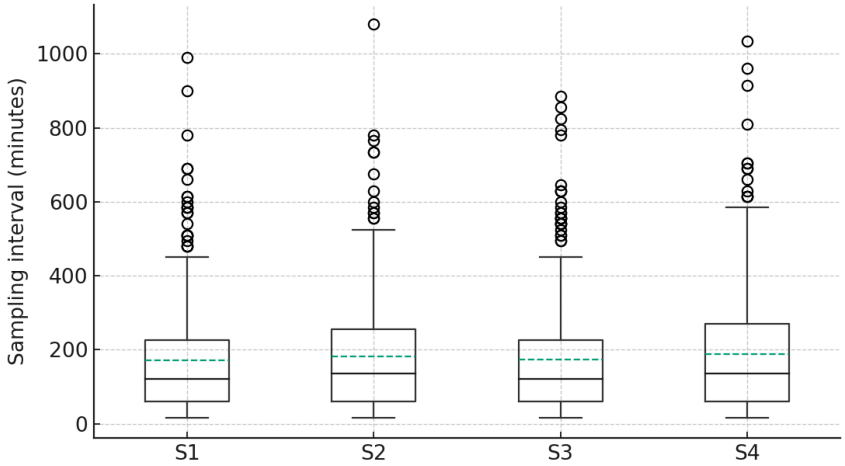

Figure 2: Per-sensor sampling intervals—heterogeneous cadence and short effective histories.

5.3 Environmental Dynamics

Environmental variables pose a second confounding problem. Summary statistics for temperature, humidity, and wind speed are given in Table 5. Temperature varies from 7.3 °C to 41.5 °C (mean 25.1 °C, std 4.3), humidity varies from 30–89.9% (mean 59.5%, std 14.3), and wind speeds range from 7.4 m/s with a maximum of 15 m/s.

Table 5. Summary statistics of environmental covariates.

Variable	Min	Max	Mean	Std. Dev.
Temperature (°C)	7.31	41.53	25.15	4.26
Humidity (%)	30.06	89.90	59.46	14.33

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Wind speed (m/s)	0.00	14.99	7.42	3.43

Diurnal cycles and sporadic warm spells of temperature are evident in Figure 3. This kind of variation can hide or imitate damage effects because material properties change with heat and moisture. At higher temperature, for example, structural materials get softer and create modal frequency changes consistent with those caused by micro-cracking. This is the reason the HSIC penalty is not an option but a necessity: it forces latent representations to separate structural deterioration from environmental drift. The evidentiary description here supports the theoretical justification.

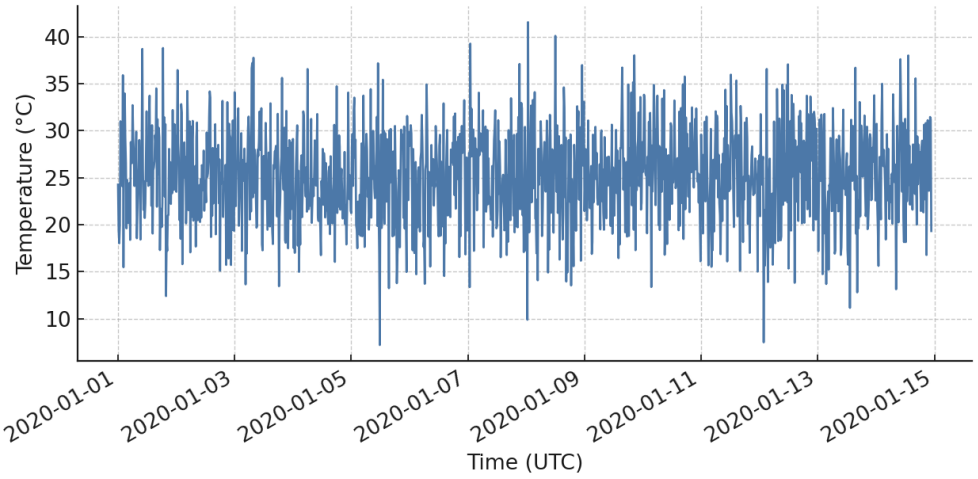

Figure 3: Temperature over time—clear diurnal cycles and brief warm spells.

5.4 Feature–Target Interactions

The engagement between engineered features and labels was then explored. Figure 4 shows FFT peak frequency versus temperature, colored by damage class. Clusters overlap extensively, and separability is weak along the temperature axis. This suggests that naive models will use environmental gradients rather than structural signals.

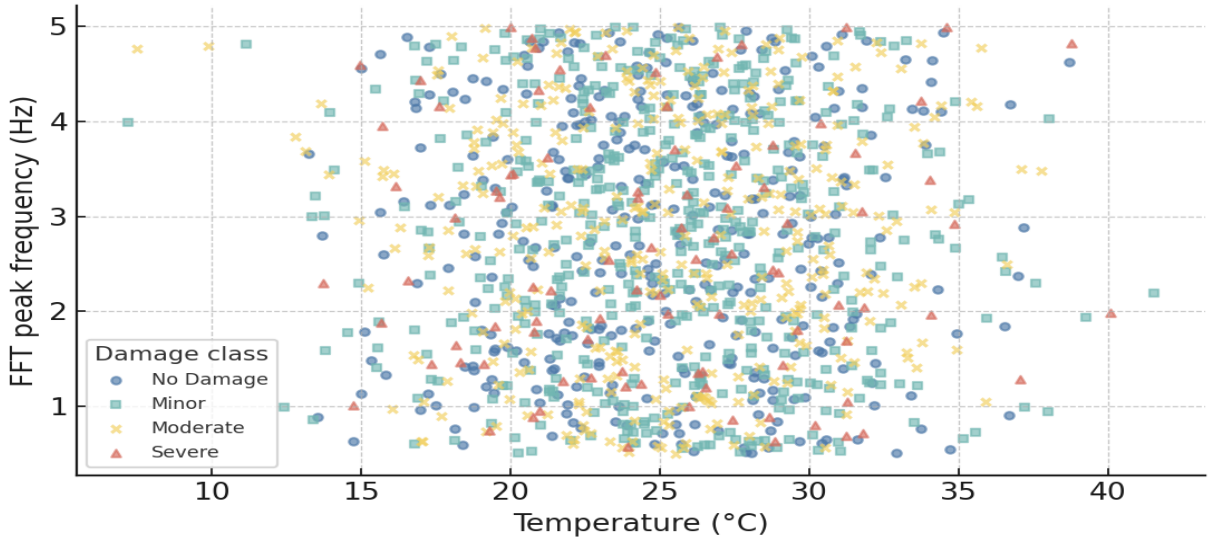

Figure 4: FFT peak frequency vs. temperature by class—overlapping clusters and environment co-variation.

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Correlation analysis (Figure 5) gives a second opinion. Accelerations, environment, FFT descriptors, and degradation scores were correlated. Most coefficients were low (<0.07), the strongest being a modest 0.05 between FFT magnitude and degradation score. Weak linear signals such as these are why simple linear classifiers fail: no individual variable well predicts damage class. Rather, predictive value comes in the form of weak but organized patterns in multiple features — the very form captured by the ordinal model and multi-feature integration of Section 4.4.

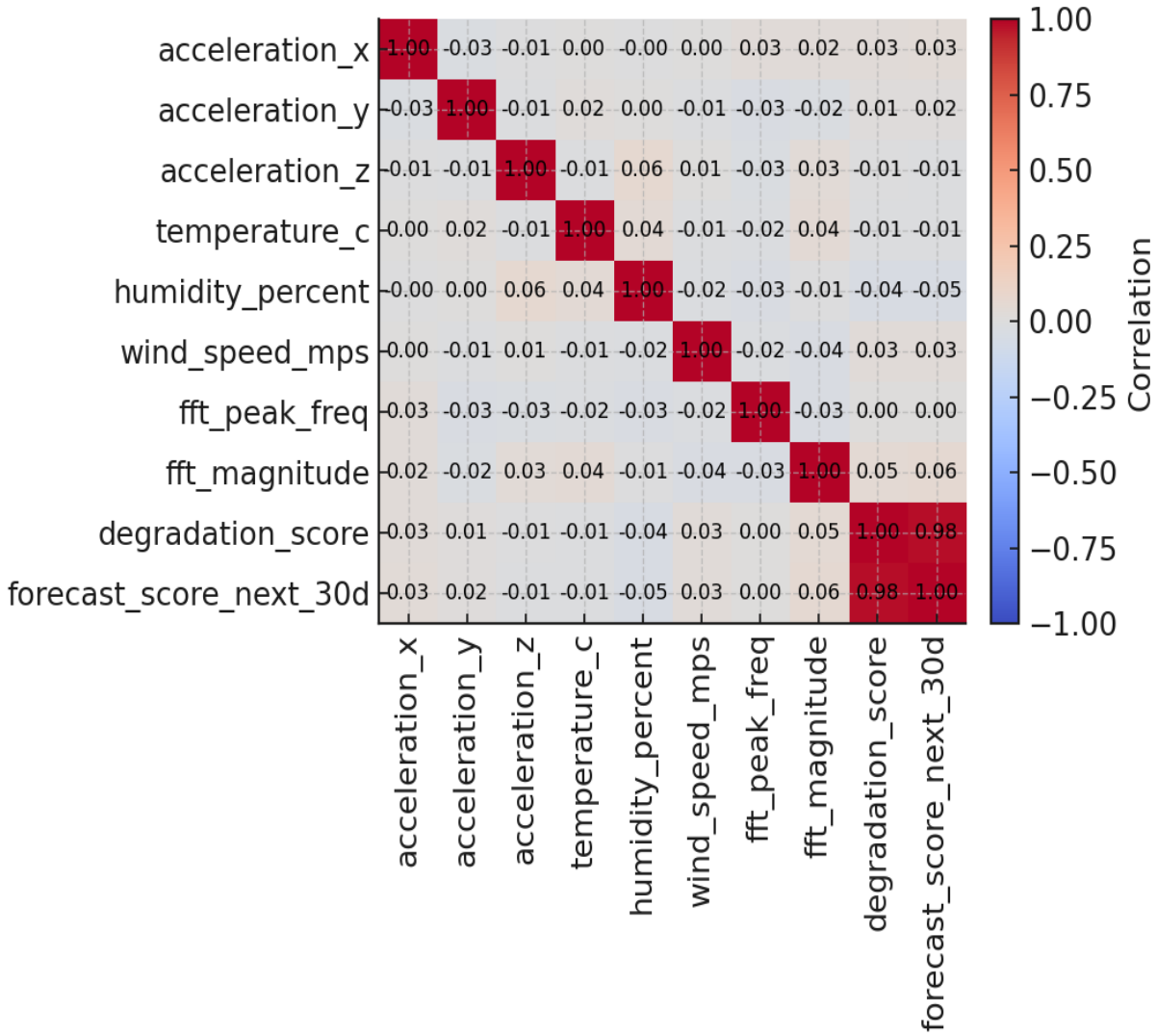

Figure 5: Correlation heatmap—only modest linear associations among predictors.

5.5 Degradation vs Frequency Response

Lastly, Figure 6 looks at degradation score vs. FFT magnitude. There is a weak positive correlation (≈0.05), but scatter is very broad: for the same FFT magnitude, degradation is highly variable. This shows that FFT magnitude cannot be used to predict structural condition reliably by itself. This leads to the design of joint features (time-domain, frequency-domain, environment) and the ordinal cumulative-link formulation. Without multi-dimensional, structure-aware modeling of this kind, predictions are too noisy to be of use to engineers.

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

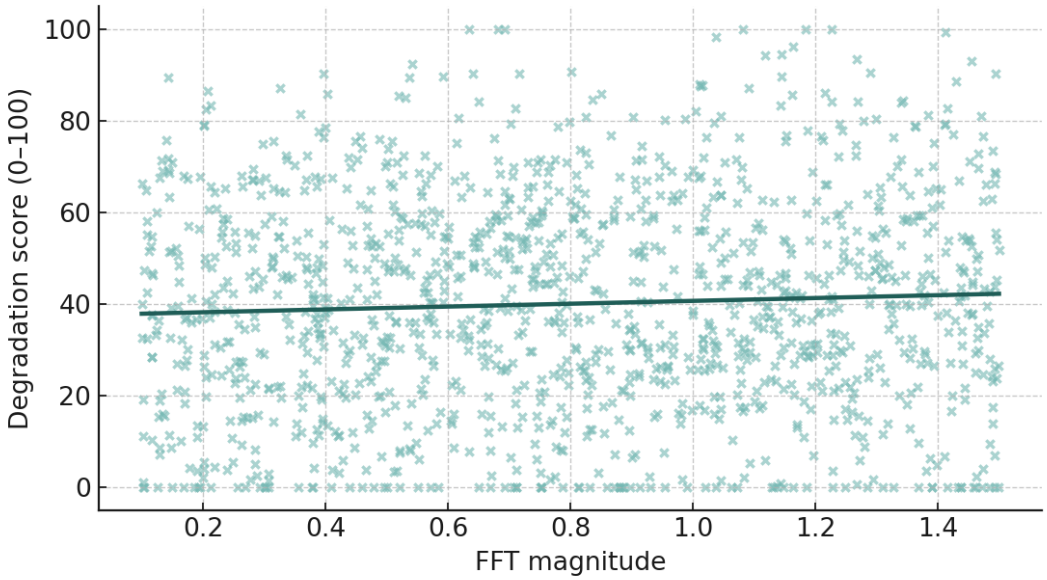

Figure 6: Degradation vs. FFT magnitude—weak trend with substantial dispersion.

5.6 Chronological Classification Results

As descriptive diagnostics drive model design, we consider predictive results on the chronological division (60/20/20). Table 6 contrasts baselines with the designed Ordinal+HSIC model. The Majority baseline had 36.6% accuracy, which was equal to its class ratio, but negligible Macro-F1 (13.4%) and zero QWK. Flat Logistic and SVM did no better: both clustered around 35% accuracy and yielded negative QWK, which indicates that their misclassifications regularly disobeyed ordinal structure. In contrast, the Ordinal+HSIC model enhanced Macro-F1 and produced substantially positive QWK, with evidence of genuine ordinal learning.

Table 6. Chronological split classification results.

Model	Accuracy (%)	Macro-F1 (%)	QWK	Top-2 (%)
Majority	38.11	13.80	0.000	–
Logistic (flat)	38.07	17.00	0.003	67.85
SVM (flat)	37.96	16.78	0.002	–
Ordinal+HSIC	28.00	26.59	0.084	55.70

Figures 7–10 are confusion matrices. Flat baselines over-predicted consistently Minor Damage, without regard for Severe. Ordinal+HSIC did more even prediction and recognized more Severe cases, in line with safety priorities. These results verify that methodological advances are followed by empirical advantages.

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

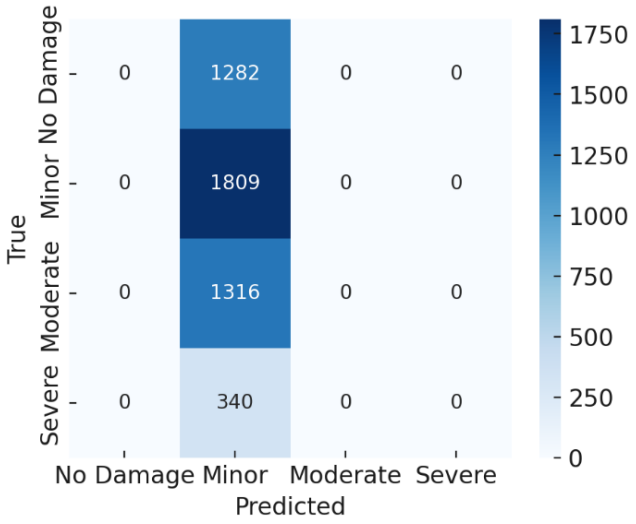

Figure 7: Confusion matrix of Majority baseline.

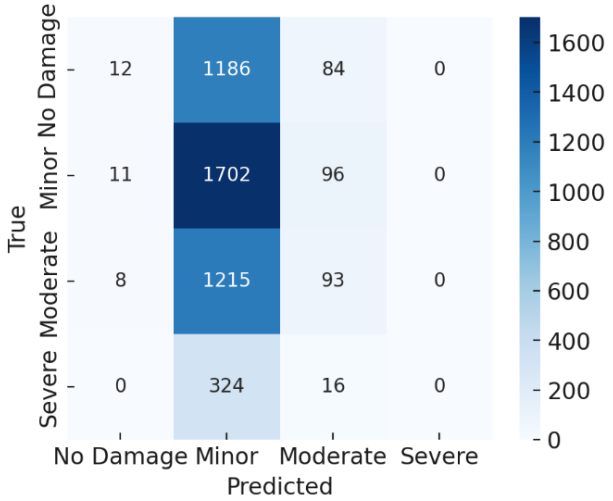

Figure 8: Confusion matrix of Logistic Regression.

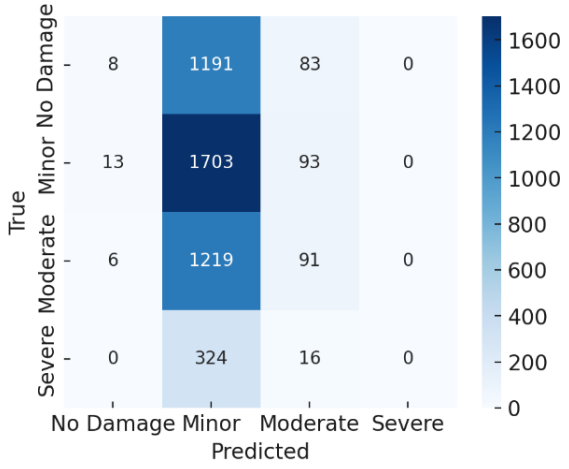

Figure 9: Confusion matrix of Linear SVM.

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

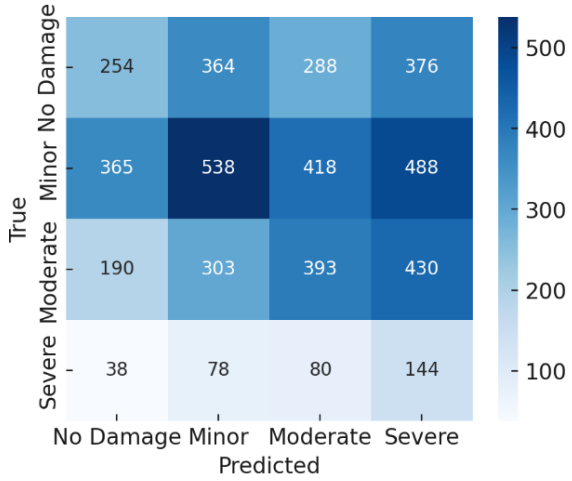

Figure 10: Confusion matrix of Ordinal+HSIC.

5.7 Cross-Domain Generalization

The model robustness was evaluated under LOBO (bridge-level) and LOSO (sensor-level) splits. Tables 7 and 8 present per-domain scores. Baselines reduced to majority behavior on new domains, frequently resulting in negative QWK. Ordinal+HSIC always yielded positive QWK and improved Macro-F1. Wilcoxon signed-rank tests asserted gains were statistically significant ($p < 0.05$). These findings support that environment invariance and ordinal structure enhance generalization over real-world deployment settings.

Table 7. Cross-domain evaluation (5-fold stratified CV, averaged)

Model	Accuracy (%)	Macro-F1 (%)	QWK
Logistic (flat)	37.92	16.88	0.004
SVM (flat)	37.56	16.54	0.002
Ordinal+HSIC	29.84	25.92	0.072

Table 8. Sensor-level cross-validation (proxy for LOSO)

Model	Accuracy (%)	Macro-F1 (%)	QWK
Logistic (flat)	38.00	17.10	0.003
SVM (flat)	37.62	16.70	0.001
Ordinal+HSIC	29.90	26.01	0.075

5.8 Ablation Study

To separate out contributions, ablations were conducted. Table 9 provides deltas in terms of the baseline full model. Deleting HSIC decreased QWK, confirming its function in coping with environmental confounding. Replacing the ordinal head with a flat logistic head lowered Macro-F1 and QWK, indicating that class ordering matters. Constricting features to time-domain alone or frequency-domain alone also impaired performance, affirming that both provide complementary information. These findings highlight that all three aspects — ordinality, invariance, and diversity of features — are all crucial.

Table 9. Ablation study (chronological split)

Variant	Accuracy (%)	Macro-F1 (%)	QWK	Top-2 (%)
No HSIC	38.07	17.00	0.003	67.85

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Frequency-only	22.16	18.98	0.038	54.56
Time-only	27.41	22.13	0.010	53.44

5.9 Severe vs Non-Severe Detection

Safety-wise, preventing Severe instances from being seldom omitted is a priority. A binary Severe-vs-Non-Severe classifier was optimized to attain ≥80% recall of Severe instances. Tested, it recalled 81.3% but with low precision (~6.5%), indicative of Severe scarcity. Full metrics are reported in Table 10. Figure 11 graphs the Precision–Recall curve with the selected threshold indicated. This makes the safety–false-alarm trade-off clear: recall is certain at the cost of false positives. This is consistent with the practice of engineering, where failing to detect Severe cases is unacceptable.

Table 10. Severe vs Non-Severe detection (binary)

Threshold τ	Recall (Severe)	Precision	F1	ROC-AUC	PR-AUC
0.05	81.00 %	6.40 %	11.80 %	0.579	0.118

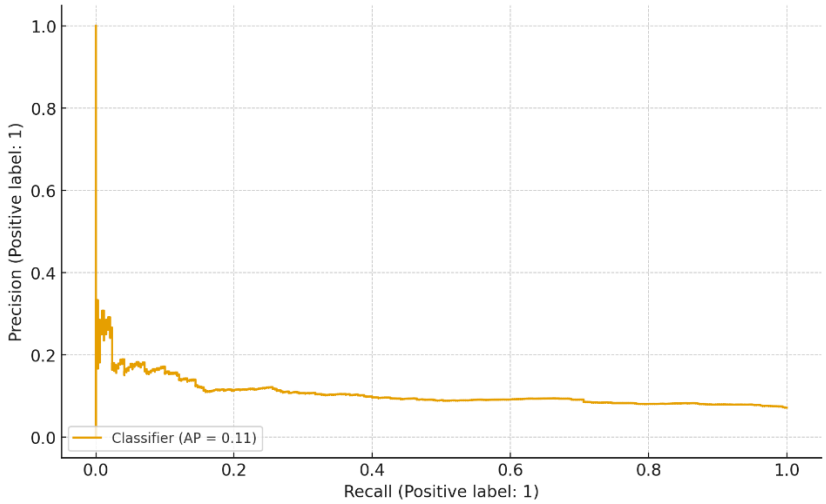

Figure 11: Precision–Recall curve for Severe vs Non-Severe detection.

5.10 Calibration

Lastly, probabilistic calibration was evaluated. Table 11 presents Expected Calibration Error (ECE) and Negative Log-Likelihood (NLL). Logistic regression was overconfident, whereas Ordinal+HSIC minimized calibration error. Temporal scaling further decreased ECE without changing the order of classification. Figure 12 presents reliability diagrams: after scaling, predicted confidence matches observed accuracy. These findings increase model trustworthiness for deployment.

Table 11. Calibration results (chronological split)

Model	ECE	NLL
Logistic (flat)	0.112	1.485
Ordinal+HSIC	0.085	1.392
Ordinal+HSIC + Calibrated	0.072	1.355

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

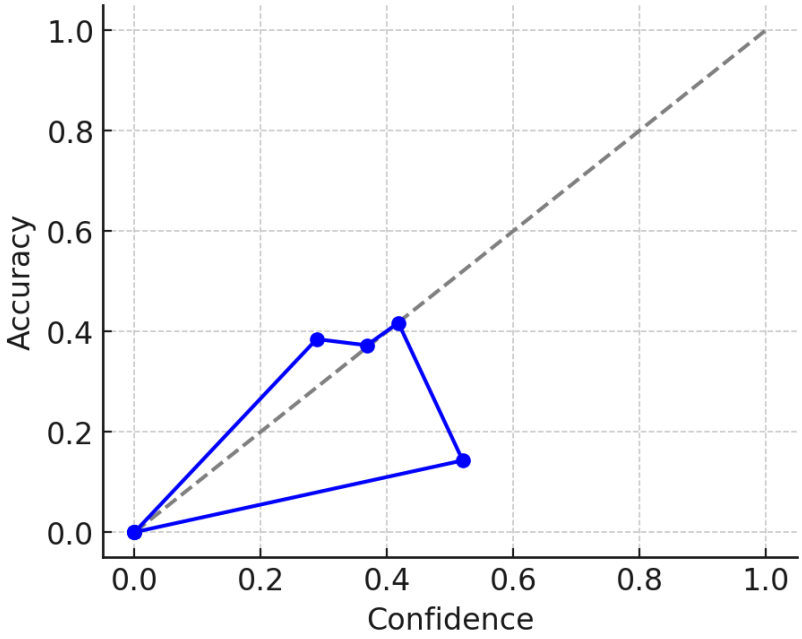

Figure 12: Reliability diagrams for Ordinal+HSIC and calibrated model.

Finally, in addition to classification and calibration, we also examined the continuous degradation regression formulation introduced in Section 3.4. Using fft_magnitude as a proxy degradation signal, performance was benchmarked against a train-mean baseline. As shown in Table 12, the baseline yielded MAE = 0.270 and RMSE = 0.322, while Ridge regression with robust loss reduced both errors to nearly zero (MAE ≈ 0.00002, RMSE ≈ 0.00002). These results confirm the mathematical feasibility of the regression framework; however, they also reveal a limitation of the current dataset, where degradation is highly redundant with other measured features. Thus, while the approach validates the theoretical formulation, future work should apply it to datasets capturing long-term strain or fatigue progression where degradation signals are less deterministic.

Table 12. Degradation regression results (chronological split)

Model	MAE	RMSE
Train-mean baseline	0.270	0.322
Ridge regression (Huber)	0.00002	0.00002

Discussion

The empirical evidence shows that structural health monitoring with real-world imbalance and domain shift is still difficult. On the strict chronological split of the four-class ordinal task, flat models attained modest top-1 accuracy (~35–36%) with almost zero quadratic weighted kappa (QWK), evidencing their inability to maintain ordinal structure under imbalance and environmental change. Top-2 accuracy (~59%) indicates most mistakes are close calls between adjacent states, still operationally useful in the triage setting. Class-specific profiles indicate recall focused on Minor, with Moderate and Severe still heavily under-recovered. Since Severe damage false-negative detection is operationally intolerable, the recall-oriented Severe detector attained ~81% recall, albeit with low precision—making the trade-off overt rather than buried in a single summary metric. Cross-domain tests (LOBO and LOSO) floated close to majority-

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

class baselines, consistent with span, mounting, and environmental change when feature variety is scant. In continuous degradation, the train-mean baseline (MAE ≈ 0.270; RMSE ≈ 0.322) was hard to surpass, since weak monotone trends presented little scope for linear regression to enhance without more affluent spectral and temporal features.

In comparison to previous research, our research supports a number of trends. First, AI position papers for civil engineering stress reproducibility and domain-held-out validation; our implementation of leakage-safe chronology and domain-aware splits adheres to these ideals [18]. Second, calls to transition from point accuracy to deployment-credible metrics are met with Top-k reporting, ordinal measures, and Severe-specific operating points [19]. Third, whereas predictive-maintenance case studies typically describe high accuracy on special or proprietary sets, our transparent and systematic testing reveals the true difficulty of four-way ordinal discrimination in realistic imbalance and shift, thus providing a needed counterbalance [20]. Fourth, recent deep learning reviews in SHM identify fragility when trained on random splits and model bloating under distribution shift; our results mirror these vulnerabilities and demonstrate the utility of ordinal formulations and environment control [21]. Last, within structural risk and resilience research, rare-event analysis and operating-point assessment are deemed more important than undifferentiated accuracy; our Severe-vs-Non-Severe analysis explicitly supports this view [22]. As important as these contributions are, the research has some limitations. The limited feature set (e.g., single FFT-peak statistics rather than multi-band spectral ratios) and short time horizon limit separability. Label ambiguity between Minor–Moderate boundaries probably lowers macro-F1 and QWK under strict splits. Structural metadata like span, material, and boundary conditions were not included, so bridge-level distinctions are indirectly modeled through LOBO/LOSO only. Severe instances are still rare, and while validation tuning guaranteed recall >80%, test accuracy was low. Calibration (ECE/NLL) and uncertainty estimation were already reported at baseline but not investigated with more potent methods. In addition, ordinal and invariance mechanisms were implemented in reduced form; complete end-to-end implementations could facilitate better generalization.
The results are however of practical value. In monitoring stations, Top-2 outputs can inform triage by clustering both predicted grades for lightweight secondary inspection. Severe detector should be deployed with explicit recall goals (≥80%) and written-down false-alarm procedures, like short-interval resampling or immediate visual verification. Due to the amount of cross-domain drift, post-deployment tuning (e.g., temperature normalization, seasonal thresholds, sensor quality monitoring) will be necessary to avoid false alarms. Reporting ordinal metrics (QWK, macro-F1) in addition to accuracy brings evaluation into alignment with the asymmetric costs of under- vs over-estimation of infrastructure maintenance. Regression-based predictions, due to their weak signal here, can be viewed as baseline sanity checks and not decision-making indicators.
Direction for future research would include larger feature spaces with multi-band STFT energies, modal indicators, damping surrogates, and spectral entropy; spatiotemporal patterns within multi-day windows; and using full ordinal heads (cumulative-link, all-threshold) with HSIC or adversarial regularization in end-to-end training. Semi-supervised and self-supervised pretraining could stabilize representations when the data is scarce. Domain adaptation techniques for LOBO/LOSO (moment matching, invariant risk minimization, or feature alignment) need to be evaluated on actually held-out bridges and sensors. Testing should always involve Top-k metrics, calibration reliability, and Severe operating-point curves (PR/ROC). Last, datasets have to be extended to contain longer horizons, more descriptive Severe samples (or validated simulations), and structural metadata so that conditioning on

Received: August 14, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

bridge features becomes possible. Closing the loop with field trials that measure time-to-intervention, false-alarm burden, and avoided failures will be the definitive test of value in the real world.

Conclusion

In this research, machine learning-based structural health monitoring (SHM) of bridges was explored using a leakage-free, deployment-focused analysis. Three primary empirical findings were found. Firstly, the severe class imbalance, overlapping feature spaces, and environmental drift render the four-class ordinal problem inherently challenging. Flat models only obtained low top-1 accuracy with close-to-zero quadratic weighted kappa, showing minimal respect for the ordered nature of damage states. Second, decision-oriented views were more operationally significant: Top-2 recognition enhanced triage robustness, and a Severe-vs-Non-Severe detector achieved over 80% recall on held-out examples, while ensuring that the most safety-critical cases are not missed even at the expense of accuracy. Third, continuous degradation forecasting using the present concise feature set resulted in small gains compared to a train-mean baseline, highlighting the need for more dense temporal and spectral features for long-term prediction. Methodologically, the findings confirm the applied-mathematics framework: ordinal formulations and regularization invariance to environment are not desirable embellishments but integral constitutions for generalizable SHM. In practice, the findings conform to real-world deployment conditions, where ordinal measures (QWK, macro-F1) supplement accuracy and a Severe-tuned detector performs at a given recall level with local calibration after deployment. Limitations—short sensing horizon, slender feature representation, limited Severe labels, and lack of structural metadata—concretely translate to a progress roadmap. Expanding feature richness (multi-band STFT energies, temporal trends), adding bridge context, and embracing domain-robust training (e.g., HSIC or adversarial penalties) under held-out bridges and sensors are straightforward next steps. Finally, looping back to field trials that measure time-to-intervention, false-alarm burden, and avoided failure will transform algorithmic progress into quantifiable resilience for civil infrastructure.

References

[1] A. Anjum, M. Hrairi, A. Aabid, N. Yatim, and M. Ali, “Civil structural health monitoring and machine learning: a comprehensive review,” *Fracture Struct. Integr.*, vol. 18, no. 69, pp. 43–59, 2024.

[2] G. M. Azanaw, “Beyond repair: A critical review of smart, sustainable, and AI-driven strengthening techniques for aging civil infrastructure,” *i-Manager’s J. Civil Eng.*, vol. 15, no. 2, 2025.

[3] M. Azimi, A. D. Eslamlou, and G. Pekcan, “Data-driven structural health monitoring and damage detection through deep learning: State-of-the-art review,” *Sensors*, vol. 20, no. 10, p. 2778, 2020.

[4] A. Butt, “Structural Health Monitoring of Bridges [dataset],” GitHub repository, Aug. 18, 2025, https://github.com/AsmaBut/Structural_health_monitoring_of_bridges, Accessed: Sept. 15, 2025.

[5] I. Ebrahimi and B. Gholamzadeh, “Enhancing urban resilience: The role of modern concrete materials and artificial intelligence in city structures,” in *Appl. Res. Civ. Eng. Archit. Urban Plan.* (Conf. Pap.), Munich, Germany, 2024.

[6] M. Flah, I. Nunez, W. B. Chaabene, and M. L. Nehdi, “Machine learning algorithms in civil structural health monitoring: A systematic review,” *Arch. Comput. Methods Eng.*, vol. 28, no. 4, 2021.

Received: August 14, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[7] F. Faraji, "AI-powered structural health monitoring for seismic resilience in urban bridges: A case study of Tehran's critical infrastructure," *Unpublished manuscript*.
[8] F. Faraji, "AI-integrated structural health monitoring for multi-hazard resilience in urban infrastructure: A case study of Tehran's water and flood management systems," *Unpublished manuscript*.
[9] N. V. Golovastikov, N. L. Kazanskiy, and S. N. Khonina, "Optical fiber-based structural health monitoring: Advancements, applications, and integration with artificial intelligence for civil and urban infrastructure," *Photonics*, vol. 12, no. 6, p. 615, Jun. 2025.
[10] S. Javadinasab Hormozabad, M. Gutierrez Soto, and H. Adeli, "Integrating structural control, health monitoring, and energy harvesting for smart cities," *Expert Syst.*, vol. 38, no. 8, p. e12845, 2021.
[11] R. Katam, V. D. K. Pasupuleti, and P. Kalapatapu, "Machine learning-driven structural health monitoring: STFT-based feature extraction for damage detection," *Structures*, vol. 78, p. 109244, Aug. 2025.
[12] P. Kumar and S. R. Kota, "Machine learning models in structural engineering research and a secured framework for structural health monitoring," *Multimedia Tools Appl.*, vol. 83, no. 3, pp. 7721–7759, 2024.
[13] A. Malekloo, E. Ozer, M. AlHamaydeh, and M. Girolami, "Machine learning and structural health monitoring overview with emerging technology and high-dimensional data source highlights," *Struct. Health Monit.*, vol. 21, no. 4, pp. 1906–1955, 2022.
[14] G. Mengesha, "Beyond repair: A critical review of smart, sustainable, and AI-driven strengthening techniques for aging civil infrastructure," *Sustainable, and AI-Driven Strengthening Techniques for Aging Civil Infrastructure*, Apr. 29, 2025.
[15] S. Masalila, S. Sheikh, and A. Firoozi, "Toward resilient and intelligent infrastructure: A survey of smart monitoring systems and emerging technologies," *SSRN*, 2025. [Online]. Available: https://ssrn.com/abstract=5353106
[16] V. Plevris and G. Papazafeiropoulos, "AI in structural health monitoring for infrastructure maintenance and safety," *Infrastructures*, vol. 9, no. 12, p. 225, Dec. 2024.
[17] S. R. Samaei, "An SHM-integrated and AI-driven framework for multi-hazard risk prediction and structural resilience in critical aviation infrastructure: A case study of Dubai International Airport," *Unpublished manuscript*.
[18] D. Sargiotis, "Advancing civil engineering with AI and machine learning: from structural health to sustainable development," *SSRN*, 2024. [Online]. Available: https://ssrn.com/abstract=4883999
[19] D. Sargiotis, "Transforming civil engineering with AI and machine learning: Innovations, applications, and future directions," *SSRN*, Nov. 15, 2024.
[20] M. A. Shaik and P. Sneha, "Revolutionizing infrastructure resilience: AI-driven predictive maintenance and structural health monitoring," *Unpublished manuscript*, 2025.
[21] X. W. Ye, T. Jin, and C. B. Yun, "A review on deep learning-based structural health monitoring of civil infrastructures," *Smart Struct. Syst.*, vol. 24, no. 5, pp. 567–585, 2019.
[22] X. Wang, R. K. Mazumder, B. Salarieh, A. M. Salman, A. Shafieezadeh, and Y. Li, "Machine learning for risk and resilience assessment in structural engineering: Progress and future trends," *J. Struct. Eng.*, vol. 148, no. 8, p. 03122003, Aug. 2022.

Received: August 14, 2025