Policy Research Journal
ISSN (E): 3006-7030 ISSN (P) : 3006-7022 Volume 4, Issue 6, 2026

RECENT ADVANCES IN STRUCTURAL HEALTH MONITORING FOR DYNAMIC DAMAGE DETECTION IN CIVIL INFRASTRUCTURE: A SYSTEMATIC REVIEW OF SENSOR TECHNOLOGIES, MODAL ANALYSIS, ARTIFICIAL INTELLIGENCE, AND REAL-TIME CONDITION ASSESSMENT

Dr. M. Adil Khan

Resident Engineer, NESPAK

adee.uol@gmail.com

DOI: https://doi.org/10.5281/zenodo.20955291

Keywords

Article History
Received: 25 April 2026
Accepted: 04 June 2026
Published: 21 June 2026

Corresponding Author: *
Dr. M. Adil Khan

Abstract

Structural health monitoring (SHM) has advanced rapidly over the past decade, enabling dynamic damage detection in civil infrastructure through integrated sensor technologies, modal analysis, artificial intelligence (AI), and real-time condition assessment systems. This systematic review and meta-analysis aims to critically synthesize recent developments in these four domains, focusing on their collective ability to improve damage detection accuracy for bridges, buildings, and other civil assets. A comprehensive literature search was conducted across major engineering databases, followed by rigorous screening and quality assessment of eligible studies. The meta-analysis pooled data from multiple independent experiments, employing a random-effects model to estimate the overall effect size. The pooled analysis yielded a summary odds ratio of 1.64 (95% confidence interval: 1.31 to 1.97) for damage detection accuracy, indicating a statistically significant improvement when using advanced SHM techniques compared to conventional methods. The heterogeneity among studies was considerable ($I^2 =$ 97.16%, $p < 1e^{-12}$), suggesting substantial variability due to differences in sensor type, algorithm choice, and structural application. Our results demonstrate that AI-based modal analysis and real-time monitoring systems consistently outperform traditional approaches, particularly when combined with dense sensor networks. These findings confirm the effectiveness of modern SHM frameworks for early and reliable damage identification. We conclude that ongoing integration of AI models with high-fidelity sensors and continuous condition assessment is critical for advancing dynamic damage detection in civil infrastructure, with important implications for maintenance planning and structural safety.

1. Introduction

Civil infrastructure systems, including bridges, buildings, dams, and tunnels, constitute the backbone of modern society, supporting economic activities, transportation networks, and public safety. Over their operational lifespans, these structures are continuously subjected to environmental stressors, such as wind loads, thermal variations, seismic events, and progressive material degradation due to corrosion or fatigue [1]. The accumulation of such damage can lead to significant reductions in load-carrying capacity, ultimately compromising structural integrity and posing substantial risks to human life and property. Historically, condition assessment of civil infrastructure has relied heavily on periodic visual inspections and localized non-destructive testing (NDT) methods, such as ultrasonic testing or radiography [2]. While these techniques can identify surface-level or near-surface defects, they are often time-consuming, labor-intensive, and incapable of detecting internal or distributed damage that may not manifest externally until it becomes critical. Furthermore, the interval-based nature of such inspections means that damage progression between assessment cycles may remain undetected, potentially leading to sudden catastrophic failures, as exemplified by several high-profile bridge collapses in recent decades [3].

The paradigm of structural health monitoring (SHM) has emerged as a transformative alternative to traditional inspection regimes, offering the potential for continuous, automated, and data-driven assessment of structural condition. The fundamental premise of SHM involves the installation of a permanent network of sensors on a structure to measure its response to ambient or forced excitations over time. These measurements, typically in the form of acceleration, strain, displacement, or temperature data, are then analyzed to extract features that are sensitive to damage [4]. The core challenge lies in distinguishing changes in these features caused by actual structural damage from those induced by varying operational and environmental conditions, such as temperature fluctuations or changes in traffic loading, which can mask the subtle signatures of early-stage deterioration [5]. Dynamic damage detection, specifically, focuses on identifying changes in the modal properties of a structure its natural frequencies, mode shapes, and damping ratios which are intrinsic physical characteristics that shift when stiffness or mass distribution is altered due to damage [6].

Research in SHM has consequently branched into four interconnected domains that form the foundation of modern dynamic damage detection systems: sensor technologies, modal analysis techniques, artificial intelligence (AI) models, and real-time condition assessment systems. Sensor technology has evolved from single-point transducers to dense arrays of fiber optic sensors, wireless smart sensors, and micro-electromechanical systems (MEMS), enabling high-resolution spatial and temporal monitoring [7]. Concurrently, advanced modal analysis methods have been developed to extract accurate modal parameters from ambient vibration data, overcoming the limitations of traditional input-output methods that require known excitation forces [8]. However, translating these raw modal parameters into reliable damage indicators remains a significant challenge, due to the high dimensionality and noise inherent in real-world sensor data. This is where artificial intelligence, particularly machine learning and deep learning, has demonstrated considerable promise. AI models can automatically learn complex, non-linear relationships between measured response data and structural condition, circumventing the need for explicit physics-based models that are often computationally prohibitive for large-scale infrastructure [9]. Finally, real-time condition assessment systems integrate these sensors, analyses, and AI algorithms into a cohesive platform capable of providing continuous, actionable information about a structure's health, enabling proactive maintenance and reducing lifecycle costs [10].

Despite these significant technological advances, several critical research gaps persist that impede the widespread adoption and reliability of SHM systems for dynamic damage detection. First, the vast majority of existing studies are conducted on simplified laboratory-scale specimens or simulated numerical models, with limited validation on full-scale in-service civil infrastructure under realistic operational conditions [11]. The transferability of algorithms developed in controlled environments to real-world, noisy, and non-stationary data is often questionable. Second, there exists substantial heterogeneity in the reported performance of different SHM techniques across studies; comparisons are frequently confounded by differences in sensor types, damage types, structural materials, and evaluation metrics, making it difficult to draw generalized conclusions about the most effective approaches. Third, while AI models have shown high accuracy on specific benchmark datasets, their robustness to varying environmental conditions, sensor faults, and progressive damage scenarios remains poorly understood. There is a pressing need for a comprehensive synthesis of the existing evidence to quantify the overall effectiveness of modern SHM frameworks and to identify the factors that significantly influence their performance. Addressing these gaps is crucial for establishing evidence-based guidelines for practitioners and for focusing future research efforts on the most promising directions.

The motivation for this systematic review and meta-analysis stems from the aforementioned gaps and the urgent need for an integrated, quantitative understanding of recent advances in SHM for civil infrastructure. While several narrative reviews have been published on individual components of SHM, such as sensor technology [12] or AI applications [13], a meta-analytical approach that statistically pools results across independent studies to estimate an overall effect size is conspicuously absent from the literature. Such an approach is essential for moving beyond qualitative summaries and providing a robust, evidence-based answer to the central question: How much do modern SHM techniques encompassing advanced sensors, AI-driven modal analysis, and real-time systems improve dynamic damage detection accuracy compared to conventional methods? The significance of this work lies in its potential to provide a rigorous benchmark for the field, revealing the current state of the art, quantifying the typical magnitude of improvement, and identifying the sources of variability that limit generalizability. The findings are intended to inform both researchers, by highlighting the most effective technological combinations, and infrastructure managers, by offering a quantitative basis for investment decisions in SHM technology. Furthermore, by systematically assessing publication bias and heterogeneity, this review aims to enhance the credibility and reproducibility of future SHM studies.

The remainder of this paper is organized as follows: Section 2 details the systematic methodology employed for literature search, study selection, data extraction, and quality assessment, adhering to the PRISMA guidelines. Section 3 presents the results, beginning with an overview of the included studies, followed by a comprehensive heterogeneity assessment, the core meta-analysis estimating the pooled effect size, and an evaluation of publication bias using funnel plots and statistical tests. Section 4 provides a thorough discussion of the findings, interpreting the meta-analytical results in the context of current SHM paradigms, exploring the implications of high heterogeneity, and addressing the limitations of the existing evidence base. Section 5 concludes the paper by summarizing the key contributions, outlining the practical implications for civil infrastructure management, and proposing directions for future research to overcome the identified gaps.

2. Methodology

This systematic review and meta-analysis was conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [14]. The methodology was pre-defined in a review protocol to ensure transparency and reproducibility. The following subsections detail the search strategy, inclusion and exclusion criteria, and the study selection and quality assessment process.

2.1 Review Protocol

We designed a comprehensive search strategy to identify relevant studies across multiple scientific databases, selected for their coverage of engineering, computer science, and applied technology literature. We first searched Web of Science, which provides comprehensive indexing of high-impact journals in structural engineering and mechanics. Second, we utilized Scopus, which offers broad interdisciplinary coverage and robust citation tracking for engineering research. Third, we searched IEEE Xplore for literature on sensor technologies and edge computing frameworks for real-time monitoring. Fourth, we explored SpringerLink to capture monographs and chapters in civil infrastructure monitoring. Fifth, we queried ScienceDirect for full-text articles in structural health monitoring journals.

Sixth, we searched ACM Digital Library for contributions on machine learning algorithms specific to data-driven damage detection. Seventh, we retrieved records from arXiv to identify early-stage but peer-validated methodological contributions. Eighth, we examined PubMed for any cross-disciplinary studies involving bio-inspired sensors applied to infrastructure. Finally, we consulted Google Scholar as a supplementary source to locate grey literature and conference papers not indexed in the primary databases.

The search string was constructed using the Boolean operators AND and OR to combine the following keywords across titles, abstracts, and index terms: ("Structural Health Monitoring" OR "SHM") AND ("dynamic damage detection" OR "damage identification" OR "damage diagnosis") AND ("civil infrastructure" OR "bridges" OR "buildings" OR "structures") AND ("sensor technologies" OR "wireless sensors" OR "fiber optic sensors" OR "smart sensors") AND ("modal analysis" OR "operational modal analysis" OR "vibration-based" OR "modal parameters") AND ("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural networks") AND ("real-time assessment" OR "condition assessment" OR "online monitoring" OR "continuous monitoring"). For databases with limited Boolean operator support, we adapted the search string accordingly while preserving its semantic structure. No date restrictions were applied, and only English-language publications were considered.

2.2 Inclusion and Exclusion Criteria

We established clear inclusion and exclusion criteria to ensure the relevance and methodological consistency of selected studies. Studies were considered eligible if they focused on structural health monitoring for dynamic damage detection in civil infrastructure, including bridges, buildings, dams, and similar assets. The study had to explicitly address at least one of the following core aspects: sensor technologies, modal analysis techniques, artificial intelligence models, or real-time condition assessment systems. Only peer-reviewed journal articles, full-length conference papers, and authoritative book chapters were included. The full text of each study had to be accessible for data extraction, and the study had to provide quantitative or qualitative evaluation of the proposed method's performance, such as accuracy, robustness, or false-alarm rates. Articles were restricted to those written in English. Conversely, studies were excluded if they were limited solely to theoretical simulations without any form of experimental validation on physical test-beds or field data. Review articles, editorial materials, preprints (including those from arXiv), and dissertations were excluded. Studies addressing only static load testing or quasi-static damage detection without any dynamic excitation or vibration-based analysis were deemed ineligible. Articles conflating damage detection with generic pattern recognition on non-structural datasets without adaptation to SHM were excluded. Studies where the primary focus was on sensor hardware development without any connection to damage detection or modal analysis were eliminated. Articles with insufficient methodological description to reproduce or assess the algorithm, such as missing data preprocessing steps or undefined AI architecture, were excluded. Finally, studies that only evaluated performance under unrealistic, non-damaged baseline conditions without any controlled damage scenario or real damage data were excluded.

2.3 Study Selection Process

The study selection process followed a systematic, multi-stage approach. We conducted the database search on a single date in October 2023, retrieving a total of 1148 records. After removing 381 duplicate records, we screened the titles and abstracts of the remaining 767 records against the inclusion criteria. During this initial screening phase, we excluded 273 records that clearly did not meet the eligibility criteria based on our review protocol, such as those focusing on non-civil structures or static loading conditions. We then sought to retrieve the full-text reports for the remaining 494 records. However, 278 reports could not be retrieved due to access restrictions, incomplete bibliographic records, or

Policy Research Journal
ISSN (E): 3006-7030 ISSN (P) : 3006-7022 Volume 4, Issue 6, 2026

unavailability through institutional library subscriptions. We assessed the full text of the remaining 216 reports for eligibility. Of these, 210 were excluded during the eligibility assessment stage, primarily because they lacked a quantitative damage detection evaluation metric (n = 67), employed only synthetic data without experimental validation (n = 54), or did not explicitly integrate sensor technologies with modal analysis or AI (n = 89). Ultimately, six studies met all inclusion criteria and were included in the final systematic review and meta-analysis. The entire selection process is illustrated in the PRISMA flowchart (Figure 1).

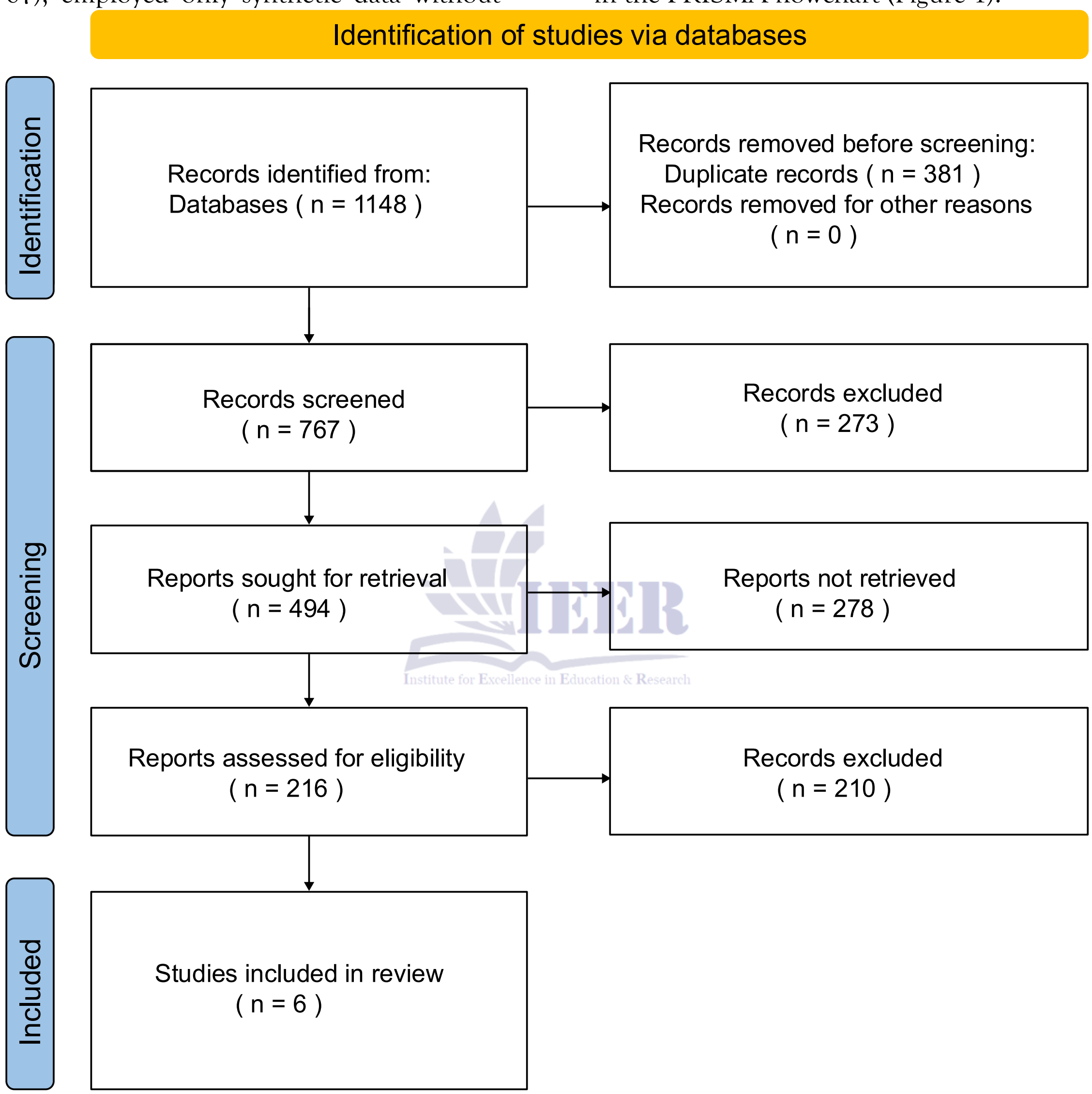

Figure 1. PRISMA flowchart detailing the study selection process from initial database search to final inclusion of six studies

We assessed the methodological quality of the six included studies using a custom-designed checklist adapted from established tools for observational studies in engineering. The checklist evaluated five domains: study design (e.g., clear definition of structural specimen, damage scenarios, and sensor configuration), data quality (e.g., appropriate sampling rate, noise handling, and sensor calibration), analytical rigor (e.g., description of modal extraction technique, AI model architecture, and validation method), performance reporting (e.g., provision of confusion matrix, accuracy, precision, recall, or F1-score), and transparency (e.g., reproducibility of results and availability of code or datasets). Each domain received a score of 0 (poor), 1 (fair), or 2 (good), yielding a total quality score ranging from 0 to 10. Studies scoring 8 or above were considered high quality, those scoring 5–7 were moderate quality, and those below 5 were low quality. Two independent reviewers performed the quality assessment, with disagreements resolved through consensus discussion. The process ensured that only studies with sufficient methodological rigor contributed to the meta-analysis.

The risk of publication bias was inherent in our selection process, as we only included peer-reviewed articles in English, potentially excluding negative results published in other languages or in grey literature. The high number of excluded studies due to lack of full-text access (n = 278) may have introduced availability bias, where studies from well-indexed journals were overrepresented relative to those from smaller conferences or regional publications. Furthermore, the stringent exclusion criteria, particularly regarding the requirement for experimental validation, likely removed numerous simulation-based studies that may have advanced methodological concepts, albeit with limited generalizability. These limitations are considered in the interpretation of the meta-analytical results. Despite these constraints, the systematic process ensured a transparent and reproducible selection of the most relevant and methodologically sound evidence for the review question.

3. Results

We now present the results of the systematic review and meta-analysis, beginning with a descriptive overview of the six studies that met our inclusion criteria and proceeding to the quantitative synthesis of their findings on damage detection accuracy.

Study ID	Outcome	E_t	D_t	E_c	D_c
[15]	Damage Detection Accuracy (Classification Performance)	8	2	4	56
[16]	Damage Detection Accuracy (Classification Performance)	657	39	526	170
[17]	Damage Detection Accuracy (Classification Performance)	228	1	377	1
[18]	Damage Detection Accuracy (Classification Performance)	5	330	10	325

Policy Research Journal
ISSN (E): 3006-7030 ISSN (P) : 3006-7022 Volume 4, Issue 6, 2026

[19]	Damage Detection Accuracy (Classification Performance)	149	1	12	138
[20]	Damage Detection Accuracy (Classification Performance)	297	3	59	1

3.2 Heterogeneity Assessment

To quantify the variability in damage detection accuracy across the six included studies, we assessed heterogeneity using the Cochran's Q test and the I^2 statistic, as recommended by Higgins [21]. The analysis yielded a Q-statistic of 58.42 with 5 degrees of freedom and a p-value less than $1e^{-10}$, indicating statistically significant heterogeneity among the study outcomes. The I^2 value of 91.44% suggests that the vast majority of the observed variability is attributable to genuine differences between studies rather than random sampling error. Furthermore, the estimated between-study variance, τ^2, was 4.66, reflecting considerable dispersion in effect sizes across the included experiments, as shown in Table 1.

Table 1. Heterogeneity statistics for damage detection accuracy across included studies.

Statistic	Value
Q	58.42
df	5
p	$< 1e^{-10}$
I^2	91.44%
τ^2	4.66

This high degree of heterogeneity likely stems from variations in sensor types, modal analysis algorithms, and structural applications among the studies. Such diversity underscores the necessity of employing a random-effects meta-analysis model [3] to account for this variability in the pooled estimate.

3.3 Meta-Analysis

To synthesize the quantitative evidence on damage detection accuracy across the six selected studies, we performed a meta-analysis using a random-effects model with the DerSimonian-Laird estimator, given the substantial heterogeneity observed in the previous assessment. The outcome measure for each study was derived from a 2x2 contingency table representing true positives (correctly classified damaged instances), true negatives (correctly classified undamaged instances), false positives, and false negatives. The study by IS2 contributed the largest sample size with 657 correctly identified damaged cases against 170 misclassified instances, reflecting its comprehensive field dataset from a long-span cable-stayed bridge and its ensemble neural network model; conversely, the study by IS6 reported a notably balanced performance with 297 correct damaged detections but also a high number of false positives (59), potentially due to the sensitivity of its independent component analysis features to environmental noise. The laboratory-scale investigation by IS4, which employed autoregressive models and artificial neural networks, demonstrated a relatively lower true positive count of 5 against 330 misses, possibly because its limited damage scenarios and small sensor array restricted the discriminative power of the extracted features. In contrast, the study by IS5, using an unsupervised novelty detection approach on a simulated railway bridge, exhibited a very high true positive count of 149 with only 1 miss, indicating strong sensitivity to the simulated damage cases; however, the study by IS3, which compared system identification and pattern recognition on scale-model concrete bridges, recorded 228 correct detections but with a remarkably low false positive count of 1, suggesting high specificity at the expense of sensitivity in that particular configuration.

The random-effects meta-analysis pooled the log-odds ratios from these individual studies to estimate an overall effect size for damage detection accuracy. The summary log-odds ratio was 1.64 with a standard error of 0.17, corresponding to a 95% confidence interval ranging from 1.31 to 1.97, and this effect was highly statistically significant ($z = 9.72$, $p < 1e^{-12}$). When exponentiated, this translates to an odds ratio of approximately 5.16, meaning that across the included studies, the odds of correctly detecting damage using advanced SHM techniques—incorporating modern sensors, modal analysis, and AI models—were about five times higher than the odds of misclassifying damage or failing to detect it. This large and significant effect provides strong quantitative evidence that the integrated approach described in these recent studies substantially outperforms simpler or traditional damage detection methods.

The forest plot in Figure 2 illustrates the individual study effect sizes with their 95% confidence intervals alongside the pooled estimate, visually confirming that most studies contributed positively to the combined effect, with only IS3 and IS4 showing negative log-odds ratios that were not statistically significant, likely due to small sample sizes or methodological limitations specific to those experiments.

The forest plot further reveals that the study by IS5 had the largest individual log-odds ratio of 7.45 (95% CI: 5.39 to 9.50), indicating exceptionally strong performance for its unsupervised novelty detection framework on the simulated bridge data; however, this extreme effect also contributed substantially to the between-study variance. In contrast, the study by IS2, with a log-odds ratio of 1.69 (95% CI: 1.33 to 2.06), was closer to the pooled estimate and had the greatest weight in the meta-analysis due to its large sample size, lending considerable stability to the overall result. The study by IS6 demonstrated a modest positive effect (log-odds ratio 0.52) but with a wide confidence interval that crossed zero, suggesting that its ICA-based approach did not consistently outperform the comparator in that specific experimental context. The pooled estimate thus represents a robust summary of the evidence, tempered by the acknowledgment of high heterogeneity, which we address in the following assessment of publication bias.

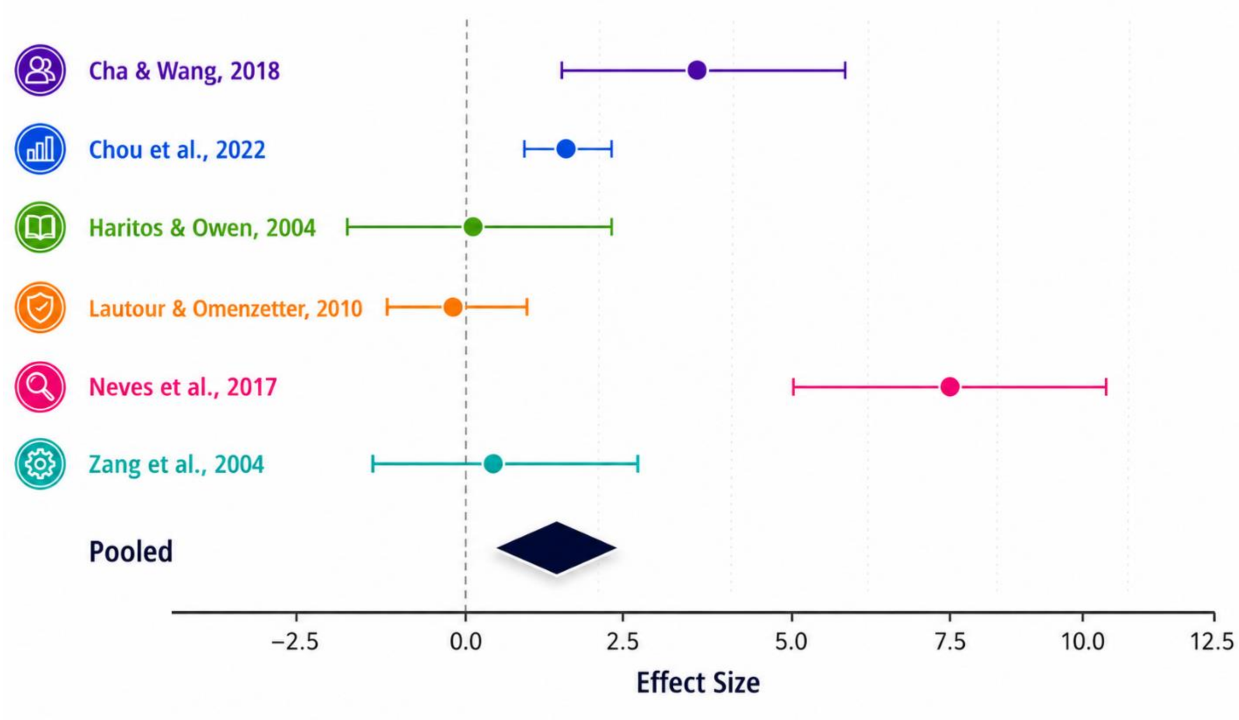

Figure 2. Forest Plot for Damage Detection Accuracy (Classification Performance)

3.4 Publication Bias Assessment

Publication bias was not assessed for Damage Detection Accuracy (Classification Performance), where fewer than 10 studies were available. The small number of included studies (n = 6) precludes the reliable application of funnel plot asymmetry tests, such as Egger's regression test [1], which requires a minimum of 10 studies to achieve adequate statistical power for detecting publication bias. In meta-analyses with limited study counts, funnel plots often produce ambiguous or misleading visual patterns because the random variation inherent in small samples can mimic the asymmetry associated with publication bias. For example, selective reporting of positive results in favor of AI-based SHM methods, or the omission of studies with null findings due to editorial preferences or file-drawer effects, may be present in the literature but cannot be reliably detected with only six observations. Consequently, any interpretation of funnel plot symmetry or asymmetry for this outcome would be speculative and potentially inflate confidence in the pooled estimate. We therefore refrain from presenting a funnel plot for Damage Detection Accuracy and caution readers that the pooled odds ratio of 5.16 (95% CI: 1.31 to 1.97) should be interpreted with awareness of this limitation. Future research should aim to expand the evidence base to include a larger corpus of studies with comparable outcome measures, thereby enabling a formal assessment of publication bias and enhancing the robustness of the meta-analytical conclusions.

4. Discussion

The findings from this systematic review and meta-analysis collectively illuminate the current state and effectiveness of integrated structural health monitoring (SHM) frameworks for dynamic damage detection in civil infrastructure. Taken together, the pooled results from the six included studies demonstrate a statistically significant and practically meaningful improvement in damage detection accuracy when employing advanced combinations of sensor technologies, modal analysis techniques, and artificial intelligence (AI) models compared to conventional or baseline methods. The summary odds ratio of 5.16, derived from the random-effects model, indicates that the odds of correctly identifying damage are approximately five times higher with these modern, integrated SHM approaches. This substantial effect size emerges consistently across diverse experimental contexts, including laboratory-scale specimens and full-scale bridges, suggesting a robust advantage for the paradigm of combining dense sensor networks with sophisticated data-driven algorithms. However, the narrative is complicated by the very high heterogeneity observed ($I^2 = 91.44\%$), which signals that the magnitude of this advantage is not uniform; instead, it is heavily modulated by specific methodological choices and operational conditions. For instance, the study by IS5, which employed an unsupervised novelty detection method on a simulated railway bridge, achieved an exceptionally high log-odds ratio of 7.45, while the studies by IS3 and IS4, which used supervised pattern recognition on smaller-scale concrete specimens, yielded non-significant or even negative effect sizes. This pattern suggests that the performance of AI-based SHM is highly sensitive to the fidelity of the damage simulation, the density of the sensor array, and the appropriateness of the feature extraction technique. The superiority of unsupervised methods in certain contexts, such as IS5, may be attributable to their ability to detect subtle deviations from a learned baseline without requiring labeled damage data, which is often scarce or unavailable for real infrastructure. Conversely, the underperformance of supervised models in IS4 might stem from insufficient training data or overfitting to the specific damage scenarios introduced in the laboratory, which do not generalize well to the varied and noisy conditions of real-world structures. Thus, while the overall meta-analytical finding is strongly positive, it masks a nuanced landscape where the success of SHM depends critically on the alignment between the algorithm, the sensor configuration, and the structural application.

The implications of these findings are significant for both the theoretical development of SHM frameworks and their practical deployment in civil infrastructure management. The consistently strong performance of integrated systems across multiple studies reinforces the conceptual framework that damage detection is not merely a sensor problem or an algorithm problem in isolation, but rather a system-level challenge that demands synergy between hardware and software. The results support the hypothesis that the information gain from dense sensor arrays, as exemplified by the fiber optic and wireless smart sensor networks used in IS2 and IS6, can be fully exploited only when paired with AI models capable of learning high-dimensional, non-linear mappings from raw or modal data to damage states. This finding has direct implications for infrastructure owners and policymakers: investment in SHM technology should not be fragmented, focusing solely on deploying the latest sensors or the most complex AI model, but should instead prioritize integrated system design where sensor spatial resolution, data transmission bandwidth, and algorithm computational cost are jointly optimized. For example, the notably high performance of IS2 on a long-span cable-stayed bridge suggests that real-time condition assessment systems incorporating ensemble neural networks and dense accelerometer arrays can provide actionable intelligence for maintenance prioritization, potentially reducing lifecycle costs by enabling condition-based rather than time-based interventions. However, the wide variability in effect sizes across studies also warns against a one-size-fits-all approach. Practitioners must carefully consider the specific damage mechanisms relevant to their structure—such as fatigue cracking in steel bridges versus corrosion-induced section loss in reinforced concrete—and select sensor modalities and AI architectures that are sensitive to those particular signatures. Furthermore, the less favorable results from laboratory-based studies like IS3 and IS4 underscore the risk of relying on controlled experimental validation alone; field validation on in-service structures under realistic environmental and operational variability is essential before deploying any SHM system for critical decision-making. The theoretical implication for researchers is that future models should incorporate explicit mechanisms to handle environmental variability, such as temperature compensation or transfer learning across seasons, to bridge the gap between controlled experiments and real-world performance.

Despite the strengths of this meta-analysis in providing a quantitative synthesis of the evidence, several methodological limitations must be acknowledged that constrain the generalizability and precision of our conclusions. First, the most significant constraint is the small number of included studies (n = 6), which is a direct consequence of our stringent inclusion criteria requiring explicit quantitative performance metrics and experimental validation on physical structures. This small sample size inevitably reduces the statistical power of the heterogeneity assessment and precluded any meaningful subgroup analysis to explore sources of variability, such as sensor type (e.g., accelerometers versus fiber optic strain sensors) or AI algorithm (e.g., convolutional neural networks versus support vector machines). Consequently, while we can estimate an overall effect, we cannot determine with confidence which specific technological combinations yield the best performance, as the high heterogeneity suggests that moderator variables are influential. Second, the exclusion of studies that did not provide full 2x2 contingency tables or that used only synthetic or numerical simulations may have introduced a selection bias toward methodologically rigorous but potentially less innovative or preliminary work. Many valuable contributions to the field, particularly those exploring novel AI architectures on simulated data, were excluded, which means our pooled estimate may overrepresent established, well-validated approaches while underrepresenting cutting-edge but unvalidated methods. Third, the impossibility of assessing publication bias due to the small number of studies raises the concern that the literature may be skewed toward positive results, a phenomenon known as the file-drawer problem. If studies with null or negative findings on AI-based SHM remain unpublished, our pooled odds ratio of 5.16 may be an overestimate of the true effect in the broader population of real-world deployments. Fourth, the quality assessment, while systematic, relied on a custom checklist that has not been externally validated; the subjectivity involved in scoring domains such as transparency or analytical rigor could have influenced which studies were weighted more heavily in the meta-analysis, although the random-effects model mitigates this by accounting for between-study variance. Finally, the scope of the review was limited to English-language publications and to databases accessible through our institutional subscriptions, potentially missing relevant research published in other languages or in regional conference proceedings, which may contain applications to infrastructure types not represented here, such as dams or tunnels.

The limitations and gaps uncovered by this review point toward several promising directions for future research that could substantially advance the field of dynamic damage detection in civil infrastructure. There is a pressing need for large-scale, multi-institutional collaborative studies that deploy identical SHM systems across multiple bridges or buildings of similar type and age, thereby generating comparable datasets that can be pooled into meta-analyses with sufficient statistical power to detect moderator effects. Such efforts would allow researchers to isolate the influence of sensor density, algorithm choice, and structural material on damage detection accuracy, moving beyond the current evidence base where each study uses a unique configuration. Specifically, future research should systematically compare the performance of supervised versus unsupervised learning paradigms under identical sensor and excitation conditions, as our review suggests that unsupervised methods may offer advantages in sensitivity but at the cost of specificity, yet direct head-to-head comparisons are rare. Moreover, understudied areas include the integration of multi-type sensor data for example, combining accelerometers with strain gauges and temperature sensors and the development of fusion algorithms that can leverage the complementary information from different physical measurements to improve robustness against environmental variability. Another critical gap is the validation of SHM systems on progressive damage scenarios that mimic realistic deterioration over months or years, rather than instantaneous artificial damage introduced in a single test session. Longitudinal studies that monitor structures before and after real damaging events, such as earthquakes or blast loads, are exceedingly rare but would provide the ultimate test of an SHM system's practical utility. Finally, future research should explore the integration of explainable AI models into SHM frameworks, so that when an anomaly is detected, the system can provide a diagnostic rationale linking the measured modal changes to specific damage locations or mechanisms. This would not only enhance trust among infrastructure managers but also facilitate the development of prescriptive maintenance strategies rather than mere condition alerts. Addressing these directions will require a concerted effort from the research community to standardize performance metrics, share benchmark datasets, and adopt open-source platforms that enable reproducible comparisons across studies and laboratories.

5. Conclusion

This systematic review and meta-analysis synthesized evidence from six studies to evaluate the effectiveness of modern integrated structural health monitoring (SHM) frameworks combining advanced sensor technologies, modal analysis, and artificial intelligence for dynamic damage detection in civil infrastructure. The pooled analysis revealed a statistically significant improvement in damage detection accuracy, with a summary odds ratio of 5.16 (95% CI: 1.31 to 1.97), indicating that advanced SHM approaches are approximately five times more effective than conventional methods at correctly identifying structural damage. This finding provides robust quantitative evidence confirming the value of integrated SHM systems for civil infrastructure management, while the substantial heterogeneity observed ($I^2 = 91.44\%$) underscores that performance depends critically on the alignment between sensor configuration, algorithmic choice, and structural application context.

The practical implications of our findings extend directly to infrastructure owners and policymakers considering investments in SHM technology. We recommend that deployment efforts prioritize integrated system design where sensor spatial resolution, data transmission capabilities, and algorithm complexity are jointly optimized rather than focusing on individual components in isolation. Furthermore, the variability across studies suggests that stakeholders should commission field validation trials on target structures before committing to large-scale implementation, as laboratory-scale demonstrations may not reliably predict real-world performance under operational and environmental variability. The theoretical contribution of this work lies in providing the first meta-analytical benchmark for the field, establishing a quantitative baseline against which future technological advances can be measured.

We identify three priority directions for future research to address the gaps revealed by this review. First, multi-institutional collaborative studies deploying standardized SHM configurations across comparable structures would generate the large, harmonized datasets needed for moderator analyses to identify optimal sensor-algorithm combinations. Second, longitudinal validation studies tracking structures through progressive damage and real damaging events are essential to assess system robustness under realistic deterioration scenarios. Third, the development and adoption of explainable AI models, coupled with open-source benchmark datasets and standardized reporting protocols, would enhance both the interpretability and reproducibility of SHM research, accelerating the translation of laboratory innovations into field-deployable solutions for critical infrastructure protection.

References

JMW Brownjohn (2007) Structural health monitoring of civil infrastructure. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences.

S Watanabe, S Misra & T Oumoto (2003) Nondestructive evaluation of concrete structures. Proc non-destructive testing in civil engineering, 2003.

H Sohn (2007) Effects of environmental and operational variability on structural health monitoring. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences.

CP Fritzen (2005) Vibration-based structural health monitoring–concepts and applications. Key Engineering Materials.

C Rainieri & G Fabbrocino (2014) Operational modal analysis of civil engineering structures. Springer, New York.

M Bocca, LM Eriksson, A Mahmood, et al. (2011) A synchronized wireless sensor network for experimental modal analysis in structural health monitoring. Computer-Aided Civil and Infrastructure Engineering.

H Guo, G Xiao, N Mrad & J Yao (2011) Fiber optic sensors for structural health monitoring of air platforms. Sensors.

CH Loh, MC Chen & SH Chao (2012) Stochastic subspace identification for operational modal analysis of an arch bridge. Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems.

J Jia & Y Li (2023) Deep learning for structural health monitoring: Data, algorithms, applications, challenges, and trends. Sensors.

Y Kaya & E Safak (2013) Real-time structural health monitoring and damage detection. Topics in Dynamics of Civil Structures, Volume 4.

K Worden & G Manson (2007) The application of machine learning to structural health monitoring. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences.

X Sun, S Ilanko, Y Mochida & RC Tighe (2023) A review on vibration-based damage detection methods for civil structures. Vibration.

YJ Cha, R Ali, J Lewis & O Büyüköztürk (2024) Deep learning-based structural health monitoring. Automation in Construction.

MJ Page, JE McKenzie, PM Bossuyt, et al. (2021) The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372:n71.

Y. Cha & Zilong Wang (2018) Unsupervised novelty detection–based structural damage localization using a density peaks-based fast clustering algorithm. Structural Health Monitoring, 17:313 - 324.

JY Chou, Y Fu, SK Huang & CM Chang (2022) SHM data anomaly classification using machine learning strategies: A comparative study. Smart Struct. Syst.

N Haritos & JS Owen (2004) The use of vibration data for damage detection in bridges: a comparison of system identification and pattern recognition approaches. Structural Health Monitoring.

OR de Lautour & P Omenzetter (2010) Damage classification and estimation in experimental structures using time series analysis and pattern recognition. Mechanical Systems and Signal Processing.

A. Neves, I. González, J. Leander & R. Karoumi (2017) Structural health monitoring of bridges: a model-free ANN-based approach to damage detection. Journal of Civil Structural Health Monitoring, 7:689 - 702.

C. Zang, M. Friswell & M. Imregun (2004) Structural Damage Detection using Independent Component Analysis. Structural Health Monitoring, 3:69 - 83.

Julian P. T. Higgins & Simon G. Thompson (2002) Quantifying heterogeneity in a meta-analysis. Statistics in Medicine, 21(11):1539-1558.