

NAVIGATING UNCERTAINTY IN IMAGE-BASED STRUCTURAL HEALTH MONITORING: THE ROLE OF FALSE POSITIVES, FALSE NEGATIVES, AND BASE RATE BIAS

Vagelis Plevris¹ and Alejandro Jiménez Ríos²

¹ Department of Civil and Environmental Engineering, College of Engineering, Qatar University
P.O. Box: 2713, Doha, Qatar
e-mail: vplevris@qu.edu.qa

² Department of Architecture & Civil Engineering, University of Bath
6 East, University of Bath, Claverton Down, Bath, BA2 7AY, UK
e-mail: ajr225@bath.ac.uk

Abstract

This study investigates the practical limitations and reliability challenges of image-based structural health monitoring (SHM) systems that utilize machine learning and computer vision to detect damage in infrastructure. While such systems offer scalable and efficient alternatives to manual inspections, their performance is significantly affected by false positives, false negatives, environmental variability, and statistical biases – particularly base rate bias. When actual damage is rare, even highly accurate models may produce an overwhelming number of false alarms, undermining trust in automated detection and leading to inefficient allocation of inspection resources. Using numerical simulations, the paper illustrates how base rate bias reduces precision in low-damage environments and explores the broader implications for system design, operational risk, and decision-making. Several mitigation strategies are presented, including threshold adjustments based on structural criticality, human-in-the-loop verification, hybrid monitoring approaches, continuous model updating, and risk-based decision frameworks. These findings emphasize the importance of context-aware deployment and uncertainty management in ensuring the effectiveness and real-world applicability of image-based SHM systems.

Keywords: Structural Health Monitoring (SHM), Image-Based Damage Detection, Base Rate Bias, Risk-Based Decision Making, Uncertainty in SHM Systems.

1 INTRODUCTION

Structural health monitoring (SHM) is essential for maintaining the safety and longevity of infrastructure systems, enabling early detection of damage and supporting timely maintenance decisions [1]. Various artificial intelligence (AI) and machine learning (ML) techniques have been successfully applied in civil engineering contexts, demonstrating their potential to revolutionize traditional inspection processes [2]. Recent studies have explored the growing influence of AI on SHM, emphasizing its role in enhancing infrastructure maintenance and safety [3-5]. Advances in ML and computer vision have led to the rise of image-based SHM techniques, which promise to transform traditional inspection processes by offering faster, scalable, and more cost-effective assessments [6-8]. Using drones, high-resolution cameras, and deep learning models, these methods automate the collection and analysis of visual data, reducing reliance on manual inspections that are often slow, labor-intensive, and subject to human variability [6, 8].

Beyond efficiency, image-based techniques provide significant practical advantages. They allow for continuous monitoring through real-time image analysis, enabling early detection of structural issues before they escalate [9, 10]. Drones and remote sensing technologies extend inspection capabilities to hard-to-reach or hazardous areas [8], improving both safety and coverage. These features make image-based methods particularly attractive for large-scale infrastructure networks, where rapid, comprehensive inspections are critical for effective asset management.

However, as image-based SHM methods move from controlled research environments to real-world applications, significant challenges have emerged. Although theoretical frameworks that account for both epistemic and aleatory uncertainties have been proposed [11, 12], practical deployments reveal vulnerabilities to false positives, where undamaged structures are mistakenly flagged as damaged, and false negatives, where actual damage goes undetected [9, 10]. Furthermore, environmental factors such as lighting variability, weather conditions, and image noise further complicate accurate damage detection [13]. These issues are not merely technical imperfections but have direct consequences for public safety, resource allocation, and decision-making.

Compounding these concerns is the often-overlooked influence of *base rate bias* [14], also known as *base rate neglect* [15]. In many practical cases, structural damage events are rare compared to the large number of inspected elements. This low base rate leads to a statistical phenomenon where even highly accurate models can produce a large number of misleading positive results, undermining the perceived reliability of automated systems [14, 16]. Without proper consideration of base rates, the outputs of sophisticated AI systems can be misinterpreted, resulting in inefficient maintenance efforts or overlooked critical risks [17].

This paper aims to explore the practical implications of uncertainty in image-based SHM by focusing on the dual impact of classification errors and statistical biases. In particular, we highlight how false positives, false negatives, and base rate bias interact to affect the trustworthiness of automated monitoring systems. Examples and numerical investigations are presented to demonstrate these effects in realistic scenarios. Strategies for mitigating uncertainty through improved system design, expert validation, and risk-based decision frameworks are also discussed, offering a practical perspective for engineers and practitioners engaged in infrastructure monitoring.

The remainder of this paper is organized as follows: Section 2 discusses key sources of uncertainty in image-based SHM systems; Section 3 explores the impact of base rate bias through illustrative examples; Section 4 presents numerical simulations; Section 5 outlines

practical strategies to mitigate uncertainty; and Section 6 concludes with a summary of findings and implications for real-world deployment.

2 TYPES OF UNCERTAINTY IN IMAGE-BASED SHM

Despite the impressive advances in ML and computer vision, image-based SHM systems remain affected by both uncertainties and classification errors. These factors can significantly impact system performance, especially when models are deployed outside controlled experimental settings [17, 18]. A key source of uncertainty is environmental variability, which may include changes in lighting, weather, or surface conditions – factors that contribute to epistemic uncertainty and can be partially reduced through better data collection and model adaptation.

In addition to such uncertainties, practical deployments are also hindered by classification errors, particularly false positives and false negatives. These outcomes result from the limitations of the model’s decision boundaries and play a critical role in shaping the reliability of damage detection. Together, uncertainty and error introduce risk to automated SHM and have important implications for both safety and maintenance efficiency.

2.1 Environmental Variability

Environmental conditions introduce a further layer of uncertainty. Changes in lighting, weather, surface moisture, and temperature can affect image quality and damage visibility, thus influencing the detection accuracy of ML models [13]. Unlike laboratory datasets, real-world conditions are highly variable, and models trained on idealized images often struggle to maintain performance in the field.

For instance, images taken under direct sunlight may exaggerate surface shadows, resembling cracks, whereas cloudy conditions may soften genuine crack patterns, making them less detectable [7, 13]. Seasonal variations, such as frost or plant growth near structures, can further complicate image interpretation and degrade detection reliability during routine monitoring.

2.2 Sensor and Acquisition-Related Uncertainty

In addition to environmental effects, image-based SHM systems are susceptible to uncertainties arising from sensor hardware and image acquisition processes. Poor camera calibration, lens distortion, and inconsistent imaging resolution can introduce systematic errors that degrade detection accuracy [19]. Occlusions – such as vehicles, vegetation, or structural components obstructing the camera’s field of view – may partially or entirely conceal damage indicators.

Furthermore, extreme or varying viewing angles can cause perspective distortion, making it difficult for models to reliably recognize features such as cracks or spalling. These acquisition-related issues are often persistent and may not fluctuate with environmental conditions, making them harder to detect and correct. Addressing this type of uncertainty requires careful sensor placement, routine calibration, and the use of preprocessing techniques that normalize or correct image quality before analysis.

2.3 False positives and false negatives

A *false positive* occurs when the system incorrectly identifies damage where none exists. While a conservative bias toward detecting potential damage may seem prudent, high false positive rates can quickly erode trust in the system. Engineers may be overwhelmed with false alarms, leading to unnecessary inspections, wasted resources, and elevated operational costs

[9]. For example, minor surface stains or harmless material discolorations may be misclassified as cracks or spalling under varying lighting conditions, triggering false alerts during routine bridge inspections [8]. False positives are particularly problematic when the cost of verification is high or when frequent false alarms cause maintenance teams to deprioritize real warnings – a phenomenon known as “alarm fatigue” [10].

In contrast, a *false negative* happens when the system fails to detect actual structural damage. False negatives pose a more serious safety risk because undetected damage may propagate, potentially leading to catastrophic structural failures. For critical infrastructure such as highway bridges or tunnels, the consequences of missing early-stage damage can be severe, both in terms of human life and economic loss [7]. Certain conditions increase the likelihood of false negatives. For instance, faint cracks may not be visible under poor lighting, fog, or oblique camera angles. In post-earthquake assessments, dust and debris may obscure true damage, causing automated image-based systems to misclassify affected regions as undamaged [6, 8].

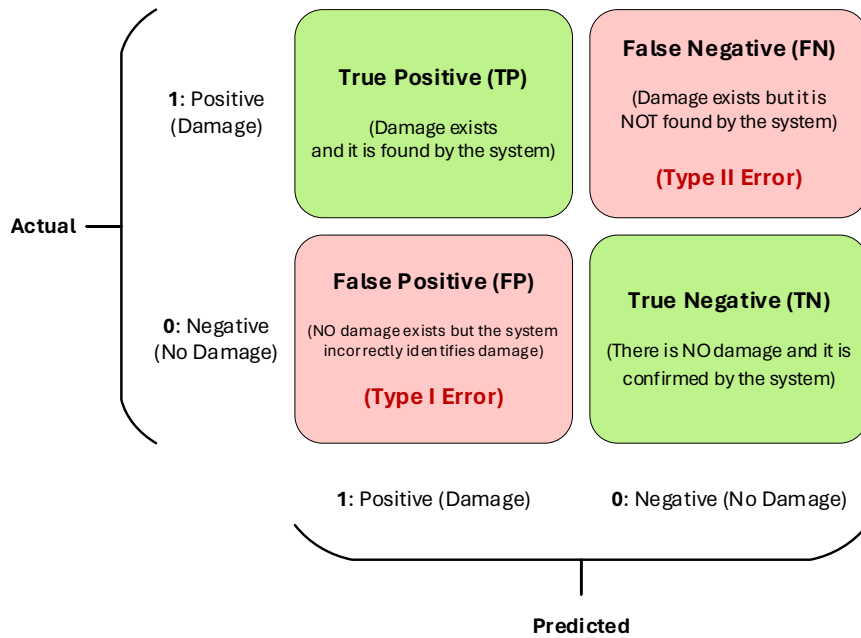


Figure 1. Confusion matrix illustrating true and false classifications in SHM damage detection.

These two types of errors are fundamental to understanding the performance of damage detection systems and are clearly illustrated in the confusion matrix shown in Figure 1, which categorizes all possible outcomes of binary classification in SHM applications.

To evaluate the performance of damage detection systems, a range of standard classification metrics can be derived from the confusion matrix shown in Figure 1. These include *Precision*, which measures the proportion of correctly identified positive cases out of all positive predictions (Eq. (1)), and *Negative predictive value*, which assesses the proportion of correctly identified negatives among all negative predictions (Eq. (2)). *Sensitivity* or *Recall* (Eq. (3)) reflects the system’s ability to correctly detect actual damage, while *Specificity* (Eq. (4)) measures the correct identification of undamaged cases. *Accuracy* (Eq. (5)) gives the overall proportion of correct classifications. Additionally, the *F1-score* (Eq. (6)) provides a harmonic mean of precision and recall, balancing both metrics into a single performance indicator. These formulas are fundamental for interpreting classification results and for comparing the performance of different SHM models in a consistent and quantitative way.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Negative Predictive Value} = \frac{TN}{TN + FN} \quad (2)$$

$$\text{Sensitivity (Recall)} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (4)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$F1\text{-score} = \frac{2}{\frac{1}{\text{Recall}} + \frac{1}{\text{Precision}}} = \frac{2 \cdot \text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}} \quad (6)$$

2.4 Combined Impact and Practical Implications

In practice, these uncertainties do not occur in isolation. A single inspection campaign may simultaneously experience environmental noise, sensor-related distortions, and produce both false positives and false negatives – while operating in a context where actual damage is rare. The combined effect can greatly diminish the value of automated assessments if not properly addressed [9].

Understanding the various sources of uncertainty is critical for designing monitoring strategies that reflect the real-world limitations of image-based systems. As discussed in the following sections, these challenges become even more pronounced when considering the role of base rate bias, where the rarity of true damage events further distorts the interpretation of positive results.

3 BASE RATE BIAS: A PRACTICAL PROBLEM

Beyond the well-recognized challenges of false positives and false negatives, a more subtle but equally critical issue in image-based SHM is the *base rate bias* [14, 15]. This statistical phenomenon can severely distort the interpretation of model outputs, even when the underlying classification models perform with high nominal accuracy. The base rate refers to the true underlying probability of an event occurring – in this case, the probability that a structure actually has damage. In most real-world monitoring scenarios, especially in routine inspections of large infrastructure networks, the occurrence of damage is relatively rare. For example, during regular bridge inspections, it is common to find that only a small fraction of inspected elements exhibit any significant structural damage [20].

When the base rate of damage is low, the performance of even highly accurate systems can become counterintuitive. Specifically, the probability that a positive detection truly indicates damage – the precision of the system – can be much lower than expected. This is because false positives, even if occurring at a low percentage, vastly outnumber true positives simply due to the sheer volume of intact structures [14, 16].

3.1 An Illustrative Example from Facial Recognition Safety Systems

To better illustrate the impact of base rate bias in detection systems, we turn to a real-world example outside of SHM: the use of facial recognition technology for security purposes. While this example touches on a sensitive application, we set aside any ethical concerns or societal controversies surrounding the deployment of such systems. Our sole intention is to present a simplified and relatable scenario that clearly demonstrates how statistical reasoning can be affected when rare events are involved.

Consider a city with a population of 1,000,000 people. Intelligence reports suggest that approximately 100 individuals in the city are classified as terrorists, representing only 0.01% of the total population. To assist in identifying these individuals, the authorities deploy an automated facial recognition system that scans people in public spaces and issues an alert when a match is found. The system's reported performance is as follows:

- It correctly identifies a terrorist 99% of the time (true positive rate, $TPR=99\%$).
- It incorrectly flags an innocent person 2% of the time (false positive rate, $FPR=2\%$).

At first glance, these figures suggest a highly reliable system. However, when examining the real-world implications, a different picture emerges. Among the 1,000,000 people:

- There are 100 actual terrorists.
- There are 999,900 innocent individuals.

Applying the system's accuracy:

- 99 of the 100 terrorists will be correctly flagged (99% true positives, $TP=99$).
- 19,998 innocent people will be incorrectly flagged as threats (2% false positives, $FP=19,998$).

Thus, the total number of individuals flagged by the system will be:

$$\text{Total flagged} = TP + FP = 99 \text{ (true positives)} + 19,998 \text{ (false positives)} = 20,097 \text{ people} \quad (7)$$

The probability that a person flagged by the system is actually a terrorist is expressed by the *Precision* value, and it can be calculated as:

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{99}{20,097} \approx 0.0049 = 0.49\% \quad (8)$$

In other words, despite the facial recognition system having an impressive individual accuracy, a person flagged by the system is, in fact, highly unlikely – less than 1 in 200 – to be a true terrorist. This example highlights a crucial lesson: when the base rate of the event being detected is extremely low, even highly accurate systems produce a disproportionate number of false positives. The overwhelming number of false alarms can undermine the practical utility of the system, leading to inefficient resource allocation, public distrust, and operational paralysis.

As we will see in the next section, similar statistical dynamics can arise in image-based SHM, particularly during routine monitoring where the occurrence of actual structural damage is rare. This example highlights the importance of considering base rates when interpreting positive detections – whether in facial recognition or SHM – because the same logic applies: when damage is rare, even accurate models can mislead decision-makers if base rate effects are ignored.

3.2 Broader Implications

The base rate bias has critical implications for infrastructure monitoring and the practical deployment of image-based SHM systems. One major consequence is its effect on resource allocation. A large number of false positives forces inspection teams to spend considerable time and financial resources verifying structures that are ultimately sound, leading to inefficiencies and diverting attention from areas that might genuinely require intervention.

Another important concern is the risk of complacency. When operators are repeatedly exposed to false alarms, they may begin to discount warnings altogether, increasing the danger of missing real signs of structural deterioration. Over time, this erosion of trust in the system's outputs can weaken the effectiveness of monitoring programs and delay necessary maintenance actions.

Finally, public safety and confidence are directly at stake. If damage detection systems are perceived as unreliable due to frequent false alerts or overlooked true positives, trust in automated infrastructure monitoring can decline. This not only undermines investment in new technologies but also creates potential risks by slowing down critical responses to actual damage.

In systems where positive findings can trigger costly inspections, evacuations, or repairs, understanding the influence of the base rate is not optional – it is essential for responsible decision-making. In the next section, we will present a series of numerical simulations that further explore how different combinations of system performance metrics and base rates affect the practical utility of image-based SHM systems.

4 NUMERICAL SIMULATIONS IN SHM

4.1 Simulation Setup

To investigate how detection performance is influenced by base rate bias and classification errors, we simulate the behavior of an image-based SHM system. This hypothetical system is designed to analyze visual data – such as photographs captured by drones or surveillance cameras – and determine whether structural damage is present in buildings or infrastructure elements. The system uses a binary classification model that labels each structure as either “damaged” or “undamaged” based on its analysis.

For all simulations, we assume a system with the following baseline characteristics unless otherwise specified:

- **True Positive Rate (TPR):** 98% (i.e., the probability of detecting damage when it exists)
- **False Positive Rate (FPR):** 5% (i.e., the probability of incorrectly flagging undamaged structures)

		Predicted		
		1. Positive (Damage)	0. Negative (No damage)	SUM
Actual	1. Positive (Damage)	<i>TPR=98%</i>	<i>FNR=2%</i>	100%
	0. Negative (No damage)	<i>FPR=5%</i>	<i>TNR=95%</i>	100%

Table 1: Normalized confusion matrix of the hypothetical SHM system, using percentage-based rates.

To explore how the system behaves under different conditions, we vary the **base rate** – the actual proportion of damaged structures in the monitored population – across seven values: 0.1%, 0.5%, 2%, 5%, 10%, 20% and 50%.

Each scenario simulates a population of 100,000 structures. For each case, we compute the number of true positives (*TP*), which represent correctly identified damaged structures, and the number of false positives (*FP*), which represent undamaged structures incorrectly flagged as damaged. Using these values, we calculate the *Precision* of the system, defined as the proportion of positive results that are correct.

4.2 Results and Observations

Table 2 summarizes the results of this simulation for a population of 100,000 structures

Base Rate (BR)	Damaged structures (DS)	Not damaged structures (NDS)	True Positives (TP=DS×TPR)	False Positives (FP=NDS×FPR)	Precision = TP/(TP+FP)
0.1%	100	99,900	98	4,995	1.92%
0.5%	500	99,500	490	4,975	8.97%
2%	2,000	98,000	1,960	4,900	28.57%
5%	5,000	95,000	4,900	4,750	50.78%
10%	10,000	90,000	9,800	4,500	68.53%
20%	20,000	80,000	19,600	4,000	83.05%
50%	50,000	50,000	49,000	4,500	95.15%

Table 2: Effect of varying base rates on true positives, false positives, and precision in a simulated SHM system monitoring 100,000 structures.

Figure 2 shows a simple plot of precision versus base rate for the assumed system parameters, where the true positive rate is fixed at 98% and the false positive rate at 5%. The graph clearly shows that precision increases significantly as the base rate of actual damage rises. When the base rate is low – as in routine monitoring scenarios – precision remains below 10%, meaning that the vast majority of positive detections are false alarms. As the base rate increases, particularly beyond 10%, the precision improves rapidly, reaching over 95% at a 50% base rate. This trend highlights the strong influence of base rate bias: even a high-performing system can produce unreliable outputs in environments where actual damage is rare. The figure underscores the need to consider the underlying base rate when interpreting detection results and evaluating the practical reliability of image-based SHM systems.

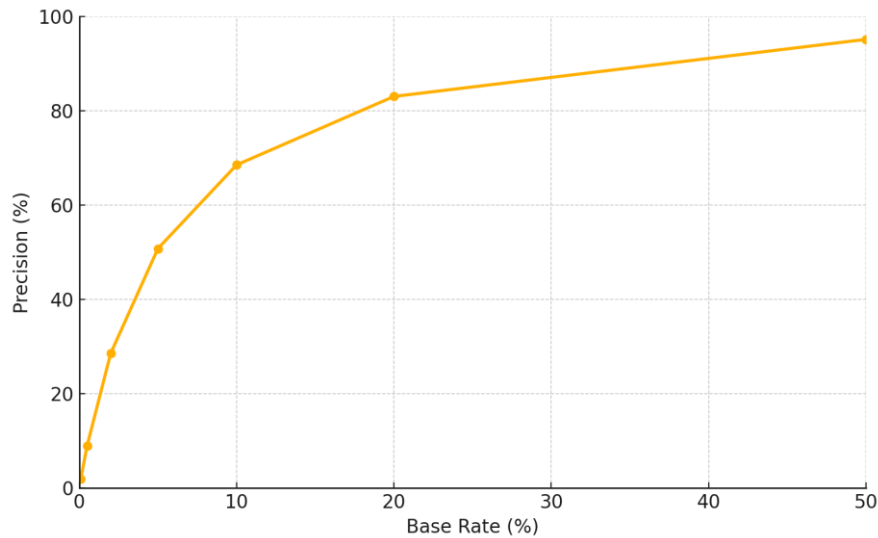


Figure 2. Precision of damage detection as a function of base rate, assuming fixed true positive (TPR=98%) and false positive (FPR=5%) rates.

4.3 Practical Takeaways

The results of the simulations reveal several important insights. First, when the base rate of actual damage is low, the impact of false positives becomes greatly amplified, making positive detections significantly less reliable. In such scenarios, even a modest false positive rate can overwhelm true detections, leading to a high number of false alarms. Second, it becomes clear that precision cannot be meaningfully interpreted in isolation; its value depends heavily on the underlying base rate of the monitored environment. Without this context, performance metrics may be misleading. Finally, for systems deployed in environments where damage is rare, achieving reliable results requires either an exceptionally low false positive rate – much lower than 5% – or the integration of additional validation mechanisms, such as human review or cross-checking with other data sources. In the next section, we examine several practical strategies designed to address these uncertainties and enhance the overall reliability of image-based SHM systems.

5 STRATEGIES TO MITIGATE UNCERTAINTY

The numerical simulations presented in the previous section clearly show that false positives, false negatives, and base rate bias can significantly impact the practical reliability of image-based SHM systems. To ensure that these systems contribute meaningfully to infrastructure management, it is essential to implement strategies that address these challenges. This section outlines practical approaches to mitigate uncertainty and improve the trustworthiness of image-based SHM in real-world applications.

5.1 Threshold Adjustment Based on Application Criticality

One effective strategy is to adjust the decision thresholds of ML models based on the criticality of the monitored structure. Many models provide a confidence score associated with each prediction. By setting a higher threshold for positive classifications, the number of false positives can be reduced – at the cost of potentially increasing false negatives.

For high-risk structures such as suspension bridges, critical tunnels, or dams, a more conservative threshold (favoring sensitivity) might be appropriate to avoid missing true damage.

Conversely, for less critical structures like minor pedestrian bridges, a higher precision threshold can help minimize unnecessary inspections.

Adaptive thresholding allows practitioners to balance risk and cost according to the specific context of each monitoring campaign.

5.2 Human-in-the-Loop Systems

Integrating human expertise into the decision-making process, commonly referred to as a human-in-the-loop (HITL) approach, offers a valuable strategy for mitigating the limitations of fully automated image-based SHM systems [21]. In this framework, positive detections generated by the system are subject to review by trained engineers before maintenance actions are taken. The human expert serves as a secondary layer of validation, using contextual judgment to confirm or override the model's predictions. This approach reduces the likelihood of false positives, strengthens confidence in system outputs, and allows for more nuanced interpretation in ambiguous cases, such as distinguishing between surface discolorations and genuine structural damage.

However, while effective in improving reliability, human-in-the-loop methods face challenges when applied at scale. For large infrastructure networks where thousands of images are processed regularly, it may be impractical to involve a human reviewer for every positive case. In such contexts, the reliance on human validation must be balanced against the need for efficiency. Selective human review, limited to high-risk detections or cases with low model confidence, may offer a compromise, preserving scalability while still addressing uncertainty in critical instances. Ultimately, the integration of human judgment should be applied strategically, complementing rather than replacing the strengths of automated SHM systems.

5.3 Hybrid Monitoring Approaches

Relying solely on image-based data can be risky, especially for detecting subtle or internal damage not visible on the surface. Combining multiple sources of information, such as vibration monitoring, strain gauges, or acoustic emission sensors, can create a hybrid monitoring system that compensates for the limitations of each method [7, 9]. For example, a positive detection from an image analysis model can be cross-validated with abnormal vibration patterns recorded by sensors. Recent work has shown that integrating artificial intelligence with traditional sensing techniques significantly enhances the robustness and responsiveness of SHM systems, particularly in the context of earthquake engineering and structural control [22]. Broader applications of AI across civil engineering, including hybrid SHM, have demonstrated the value of combining intelligent data analysis with diverse sensing strategies to improve damage detection and overall system reliability [23]. Such fusion of data streams improves overall confidence and reduces the probability of both false positives and false negatives.

5.4 Continuous Model Updating and Retraining

Image-based SHM systems should not be static. Regular retraining of ML models using updated datasets, especially with new examples of both damage and non-damage under diverse environmental conditions, can significantly improve performance over time [24]. Including real-world false positives and false negatives encountered during previous inspections into the training process allows the system to learn from its past mistakes. Adaptive models that evolve with time are better suited to cope with the variability and complexity of real-world environments.

5.5 Risk-Based Decision Frameworks

Ultimately, uncertainty in SHM should be addressed within a risk-based decision-making framework. Rather than treating all positive detections as equally important, decision-makers can assign varying levels of response based on several factors, including the criticality of the structure, the severity and confidence level associated with the detected damage, and the potential consequences of failing to identify actual damage. For instance, a low-confidence detection in a minor structure may only warrant a scheduled follow-up inspection, whereas the same detection in a major highway bridge could justify immediate on-site verification. This kind of tiered response strategy allows for more efficient use of resources by ensuring that the highest level of attention is directed toward the most critical and high-risk situations [25-27].

6 CONCLUSIONS

Image-based SHM systems, powered by advances in machine learning and computer vision, offer significant potential for enhancing the efficiency, scalability, and reach of infrastructure inspection and maintenance. However, this study demonstrates that practical deployment of such systems involves far more complexity than algorithmic accuracy alone can address. The challenges posed by false positives, false negatives, and particularly base rate bias are substantial and cannot be ignored in real-world applications.

Our simulations and examples show that even high-performing models can yield a large number of false alarms when actual damage is rare – leading to wasted resources, unnecessary inspections, and reduced confidence in system outputs. Conversely, overlooking false negatives can compromise structural safety, especially in critical infrastructure. These findings highlight the importance of understanding uncertainty not as an exception but as a core feature of SHM system design and interpretation.

To manage these limitations, SHM systems must be adapted to their operational context. This includes tuning thresholds based on structure criticality, incorporating human-in-the-loop validation, and leveraging additional data sources for cross-verification. No single configuration or approach fits all scenarios. Instead, systems must be context-aware and capable of balancing risk, cost, and reliability.

Rather than replacing manual inspection methods outright, image-based SHM should be viewed as a decision-support tool – one that assists engineers and asset managers in making informed, risk-based decisions. Effective deployment requires smarter integration of technology with human oversight, especially in low-damage environments where false alarms can erode trust.

Future work should focus on improving system robustness to environmental variability, expanding hybrid and multimodal monitoring solutions, and enhancing the interpretability of detection results through clear communication of uncertainty. By embracing the complexities of uncertainty and designing systems around them, image-based SHM can play a transformative role in supporting safer, more resilient, and more cost-effective infrastructure management.

REFERENCES

- [1] Wang, G. and J. Ke, *Literature Review on the Structural Health Monitoring (SHM) of Sustainable Civil Infrastructure: An Analysis of Influencing Factors in the Implementation*. Buildings, 2024. **14**(2): p. 402 DOI: <https://doi.org/10.3390/buildings14020402>.
- [2] Plevris, V., A. Ahmad, and N.D. Lagaros, eds. *Artificial Intelligence and Machine Learning Techniques for Civil Engineering*. 2023, IGI Global. DOI: <https://doi.org/10.4018/978-1-6684-5643-9>.
- [3] Plevris, V. and G. Papazafeiropoulos, *AI in Structural Health Monitoring for Infrastructure Maintenance and Safety*. Infrastructures, 2024. **9**(12) DOI: <https://doi.org/10.3390/infrastructures9120225>.
- [4] Plevris, V., *AI-Driven Innovations in Earthquake Risk Mitigation: A Future-Focused Perspective*. Geosciences, 2024. **14**(9): p. 244 DOI: <https://doi.org/10.3390/geosciences14090244>.
- [5] Jiménez Rios, A., V. Plevris, and M. Nogal, *Bridge management through digital twin-based anomaly detection systems: A systematic review*. Frontiers in Built Environment, 2023. **9**(1176621) DOI: <https://doi.org/10.3389/fbuil.2023.1176621>.
- [6] Kim, J.-W., H.-W. Choi, S.-K. Kim, and W.S. Na, *Review of Image-Processing-Based Technology for Structural Health Monitoring of Civil Infrastructures*. Journal of Imaging, 2024. **10**(4): p. 93 DOI: <https://doi.org/10.3390/jimaging10040093>.
- [7] Spencer, B.F., V. Hoskere, and Y. Narazaki, *Advances in Computer Vision-Based Civil Infrastructure Inspection and Monitoring*. Engineering, 2019. **5**(2): p. 199-222 DOI: <https://doi.org/https://doi.org/10.1016/j.eng.2018.11.030>.
- [8] Mandirola, M., et al., *Use of UAS for damage inspection and assessment of bridge infrastructures*. International Journal of Disaster Risk Reduction, 2022. **72**: p. 102824 DOI: <https://doi.org/10.1016/j.ijdrr.2022.102824>.
- [9] Fan, C.-L., *Deep neural networks for automated damage classification in image-based visual data of reinforced concrete structures*. Heliyon, 2024. **10**(19): p. e38104 DOI: <https://doi.org/10.1016/j.heliyon.2024.e38104>.
- [10] Poudel, U.P., G. Fu, and J. Ye, *Structural damage detection using digital video imaging technique and wavelet transformation*. Journal of Sound and Vibration, 2005. **286**(4): p. 869-895 DOI: <https://doi.org/10.1016/j.jsv.2004.10.043>.
- [11] Jiménez Rios, A., V. Plevris, and M. Nogal, *Synthetic Data Generation for the Creation of Bridge Digital Twins What-if Scenarios*, in *9th International Conference on Computational Methods in Structural Dynamics and Earthquake Engineering (COMPdyn 2023)*. 2023: Athens, Greece. p. 4801-4809. DOI: <https://doi.org/10.7712/120123.10760.21262>.
- [12] Jiménez Rios, A., V. Plevris, and M. Nogal, *Uncertainties in the Synthetic Data Generation for the Creation of Bridge Digital Twins*, in *5th International Conference on Uncertainty Quantification in Computational Science and Engineering (UNCECOMP 2023)*. 2023: Athens, Greece. p. 39-47. DOI: <https://doi.org/10.7712/120223.10323.20020>.
- [13] Torzoni, M., L. Rosafalco, A. Manzoni, S. Mariani, and A. Corigliano, *SHM under varying environmental conditions: an approach based on model order reduction and deep learning*. Computers & Structures, 2022. **266**: p. 106790 DOI: <https://doi.org/10.1016/j.compstruc.2022.106790>.
- [14] Bar-Hillel, M., *The base-rate fallacy in probability judgments*. Acta Psychologica, 1980. **44**(3): p. 211-233 DOI: [https://doi.org/10.1016/0001-6918\(80\)90046-3](https://doi.org/10.1016/0001-6918(80)90046-3).

- [15] Stengård, E., P. Juslin, U. Hahn, and R. van den Berg, *On the generality and cognitive basis of base-rate neglect*. Cognition, 2022. **226**: p. 105160 DOI: <https://doi.org/10.1016/j.cognition.2022.105160>.
- [16] Welsh, M.B. and D.J. Navarro, *Seeing is believing: Priors, trust, and base rate neglect*. Organizational Behavior and Human Decision Processes, 2012. **119**(1): p. 1-14 DOI: <https://doi.org/10.1016/j.obhdp.2012.04.001>.
- [17] Plevris, V., *Assessing uncertainty in image-based monitoring: addressing false positives, false negatives, and base rate bias in structural health evaluation*. Stochastic Environmental Research and Risk Assessment, 2025. **39**: p. 959-972 DOI: <https://doi.org/10.1007/s00477-024-02898-7>.
- [18] Plevris, V., *Addressing the Pitfalls of Image-Based Structural Health Monitoring: A Focus on False Positives, False Negatives, and Base Rate Bias*. ArXiv e-prints, 2024(arXiv:2410.20384) DOI: <https://doi.org/10.48550/arXiv.2410.20384>.
- [19] Yin, W., X. Zang, L. Wu, X. Zhang, and J. Zhao, *A Distortion Correction Method Based on Actual Camera Imaging Principles*. Sensors, 2024. **24**(8): p. 2406 DOI: <https://doi.org/10.3390/s24082406>.
- [20] Fard, F. and F. Sadeghi Naieni Fard, *Development and Utilization of Bridge Data of the United States for Predicting Deck Condition Rating Using Random Forest, XGBoost, and Artificial Neural Network*. Remote Sensing, 2024. **16**(2): p. 367 DOI: <https://doi.org/10.3390/rs16020367>.
- [21] Mosqueira-Rey, E., E. Hernández-Pereira, D. Alonso-Ríos, J. Bobes-Bascarán, and Á. Fernández-Leal, *Human-in-the-loop machine learning: a state of the art*. Artificial Intelligence Review, 2023. **56**(4): p. 3005-3054 DOI: <https://doi.org/10.1007/s10462-022-10246-w>.
- [22] Lu, X., V. Plevris, G. Tsiasas, and D. De Domenico, *Editorial: Artificial Intelligence-Powered Methodologies and Applications in Earthquake and Structural Engineering*. Frontiers in Built Environment, 2022. **8** DOI: <https://doi.org/10.3389/fbuil.2022.876077>.
- [23] Lagaros, N.D. and V. Plevris, *Artificial Intelligence (AI) Applied in Civil Engineering*. Applied Sciences, 2022. **12**(15) DOI: <https://doi.org/10.3390/app12157595>.
- [24] Alzubaidi, L., et al., *A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications*. Journal of Big Data, 2023. **10**(1): p. 46 DOI: <https://doi.org/10.1186/s40537-023-00727-2>.
- [25] Hughes, A.J., R.J. Barthorpe, N. Dervilis, C.R. Farrar, and K. Worden, *A probabilistic risk-based decision framework for structural health monitoring*. Mechanical Systems and Signal Processing, 2021. **150**: p. 107339 DOI: <https://doi.org/10.1016/j.ymssp.2020.107339>.
- [26] Feghhi, M., S.S. Khedmatgozar Dolati, A. Mehrabi, P. Malla, and D. Tabiatnejad, *A New Framework for Condition and Risk Assessment for Sustainable Management of PT Bridges*. Sustainability, 2024. **16**(22): p. 9703 DOI: <https://doi.org/10.3390/su16229703>.
- [27] Hughes, A.J., et al., *On risk-based active learning for structural health monitoring*. Mechanical Systems and Signal Processing, 2022. **167**: p. 108569 DOI: <https://doi.org/10.1016/j.ymssp.2021.108569>.