Structural Key Performance Indicators for Condition Monitoring of Concrete Bridges Using Artificial Intelligence: A Review

Mohammadjavad Berangi[1], Fengqiao Zhang*[,1], Wassamon Phusakulkajorn[1,2], Alfredo Núñez [1], Kumar Anupam[1],

[1]*Department of Engineering Structures, Delft University of Technology, Delft, 2628 CD, the Netherlands*

[2] *Rail and Modern Transport Research Group, National Metal and Materials Technology Center (MTEC), Pathum Thani, Thailand*

*Corresponding author: f.zhang-5@tudelft.nl

ARTICLE INFO

ABSTRACT

Keywords
Artificial Intelligence,
Structural Health Monitoring,
Key Performance Indicators,
Structural Assessment,
Concrete Bridges,

The actual service life of concrete bridges often deviates from design expectations due to varied loading histories and exposure to environmental and traffic-induced deterioration. Accurately assessing these deviations is essential yet remains challenging. Structural health monitoring (SHM) systems offer a pathway to evaluate bridge conditions and guide maintenance decisions, but their adoption is limited, and the data they generate is often underutilized. Recent advances in artificial intelligence (AI) present new opportunities to enhance SHM by extracting actionable insights from complex datasets. This paper presents a structured review of AI applications in the assessment of real concrete bridges, emphasizing approaches that derive structural key performance indicators (sKPIs). Methods are categorized by input data type and assessed for their performance in damage detection, capacity estimation, multi-type damage classification, signal decomposition, etc. Key challenges are discussed, including data scarcity, interpretability, and robustness. The review concludes with recommendations to advance AI toward practical, scalable implementation in bridge assessment.

1 Introduction

Concrete has been widely used in bridge construction due to its robust mechanical properties and cost-effectiveness [1]. Globally, a significant number of concrete bridges were constructed in the mid-20th century [2], and in the Netherlands, nearly half of them date from this period [3]. Designed for a service life of 50 to 100 years, many of these bridges are now reaching the end of their expected lifespan. Material degradation, increased traffic loads, and outdated design codes compromise their compliance with current safety standards [4-6]. In the context described above, continuous structural health monitoring (SHM) and timely maintenance are critical to ensure public safety and minimize the economic and environmental impacts of bridge failures.

SHM provides information on material properties (e.g., elastic modulus), localized defects (e.g., cracks, corrosion, delamination), and structural dynamic performance (e.g., vibrations and deflections) [7]. This information can help estimate structural capacity by using suitable and tailored structural models, which can be empirical, analytical, or numerical [8]. Numerical modeling often requires comprehensive structural information, such as geometry, material properties, and localized damage, which may not always be available or can be computationally intensive. Alternatively, analytical or empirical models can link the structural capacity to structural key performance indicators (sKPIs). The sKPIs are typically derived from measurable parameters such as crack location and size, stress/strain at critical cross sections, or structural vibration and deflection [9].

Various SHM techniques are available to measure the above-mentioned parameters [7, 10]. Crack measurement or detection can be carried out using optical techniques such as digital image correlation (DIC), infrared thermography (IR), and fiber optic sensors (FOS), or vibration-based techniques like ultrasonic pulse velocity (UPV), acoustic emission (AE), and impact echo (IE). Strain and stress are typically measured using linear variable differential transformers (LVDTs), strain gauges, or FOS. While structural vibrations are traditionally monitored using accelerometers, more advanced systems such as laser doppler vibrometers (LDVs) offer high resolution. The data collected from the different SHM monitoring techniques vary in formats, frequencies, and qualities. Furthermore, the extensive and ongoing application of these techniques generates substantial data volumes, creating issues for efficient storage, processing, and analysis. [11].

Given the scale and complexity of concrete bridge monitoring data, traditional analysis models often prove inefficient. The inefficiency can stem from the reliance on simplified models, the difficulty of integrating diverse sensor data, and the computational demands of processing large datasets under operational conditions. To address these issues, artificial intelligence (AI) has emerged as a promising solution, offering advanced capabilities that have been applied in various fields. AI refers to computational systems that can simulate human intelligence to perform different tasks, such as data analysis and pattern recognition [12]. Recent advancements in AI have enabled its growing application in bridge monitoring, particularly in processing large datasets [13]. Specifically, AI can enhance the extraction of sKPIs, enabling the detection and classification of different types of bridge damage, such as cracks, spalling, and reinforcement buckling [14]. Furthermore, AI can assist in predicting trends, such as the degradation of the structural capacity of the bridge over time [15].

Despite the growing interest in applying AI to the SHM of concrete bridges, its practical deployment remains limited. A major barrier is that bridges are often not instrumented with sufficient sensors to enable continuous inspection. When monitoring data is sparse due to limited instrumentation or infrequent inspections, the effectiveness and accuracy of AI-based methods can be reduced. Besides, AI models, particularly deep learning algorithms, are ‘black boxes’, meaning the rationale behind the predictions can be opaque and difficult to interpret. The lack of interpretability poses a serious concern, especially in safety-critical tasks like structural maintenance, where transparency and accountability are essential [16]. These tasks often rely on detailed knowledge of material properties, load distributions, and dynamic responses, which are more effectively addressed using physics-based models. To alleviate the above-mentioned issue, the integration of physics-based knowledge into AI, known as physics-informed machine learning (ML), has emerged as a promising approach. By embedding domain-specific physical laws into data-driven models, this hybrid method can improve both accuracy and reliability [17, 18].

Considering the issues discussed above, the suitability of AI should be carefully assessed for each use case. Key considerations include the availability and quality of monitoring data, the extent to which physical understanding is required, the ability of the models to generalize across different structures, and the need for interpretability in decision-making processes. Several prior reviews [11, 13, 19, 20] provided comprehensive overviews of AI models in bridge SHM and have advanced the understanding of the models' capabilities and their application contexts. However, the issues with these reviews are that they often do not link AI techniques to a structured set of **sKPIs** or offer a task-oriented comparison of model suitability across diverse SHM data types. To address these issues, the present study aims to:

- Structure the review around sKPIs that are directly relevant to concrete bridge assessment,
- Provide a comparative evaluation of AI models for specific SHM tasks, detailing strengths, limitations, and task–data suitability,
- Offer practical directions for future research and deployment to align AI-based SHM with engineering decision-making for concrete bridge management.

To address the aim, the remainder of this paper is organized as follows. To address the aim of this study, the remainder of the paper is structured as follows. Section 2 presents the methodology adopted for retrieving relevant articles. An overview of the retrieved and analyzed studies is then provided in Section 3. The identification of sKPIs for the assessment of concrete bridges, together with the monitoring techniques that provide the related data, is discussed in Section 4. Section 5 offers a review of AI methods that can be employed to relate raw monitoring data with the identified sKPIs. The performance of AI methods across different application cases is examined in Section 6. Then, potential future applications and advancements of AI in the structural assessment of concrete bridges are explored in Section 7. The paper concludes in Section 8 with a synthesis of key insights and closing reflections on the role of AI in the structural assessment of concrete bridges.

2 Methodology

While AI has been widely applied to other contexts such as metallic structures, buildings, and laboratory-based experiments, this review concentrates explicitly on its use in real-world monitoring of concrete bridges. The reason for such concentration is that monitoring real concrete bridges presents a distinct set of challenges that are not often encountered in other contexts. Challenges like the heterogeneous and complex nature of concrete as a material, the large spatial extent that must be monitored, varying and often harsh environmental conditions, and the presence of pre-existing damage accumulated over years of service. Therefore, when applying AI to process monitoring data, these unique challenges should be considered.

The review process in the study is carried out by following a structured approach inspired by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [21]. It is noted that while this review is not a full systematic review, key principles such as inclusion and exclusion criteria from the PRISMA framework were adopted to enhance the clarity and reproducibility of the literature selection and analysis process. Table 1 summarizes the literature search process and selection criteria employed in this review. As shown in the corresponding table, the process involved retrieving peer-reviewed journal articles from the Scopus database using a structured query, followed by the inclusion and exclusion criteria to ensure relevance to real-world AI applications in the SHM of concrete bridges.

In the SHM of concrete bridges, monitoring and inspection are often regarded as complementary and inseparable components. Monitoring typically refers to the long-term, continuous evaluation of structural performance, generating data streams that capture the temporal evolution of sKPIs [7]. Inspection, by contrast, is performed periodically, often through visual surveys, scans, or non-destructive tests, and provides discrete snapshots of the structural condition [10]. While both approaches address similar sKPI-related parameters (e.g., crack width, deflection, corrosion), they differ in temporal resolution and practical constraints. In this paper, the term "*monitoring data*" is used in a broad sense to encompass both monitoring and inspection activities, consistent with the scope of the literature search in Table 1(including monitoring, detection, or diagnosis). The AI techniques reviewed in this paper are applicable to both, though computational efficiency and cost-effectiveness should be considered in the monitoring applications to enable near real-time performance.

Table 1.Literature Search and Selection Criteria

Criteria	Explanation
Database	Scopus was selected for its broad coverage of peer-reviewed literature in the engineering domain.
Time Frame	Publications between 2001 and 2024 were considered.
Language	Only publications written in English were included.
Document Type	Only journal articles were included; conference papers, reviews, and non-peer-reviewed documents were excluded.
Subject Area	Publications were restricted to the Engineering subject domain.
Core Query	TITLE-ABS-KEY((computer) OR (machine AND learning) OR (artificial AND intelligence*) OR (big AND data)) AND TITLE-ABS-KEY((concrete AND bridge*)) AND TITLE-ABS-KEY(monitoring OR detection OR diagnosis*) AND PUBYEAR > 2000 AND PUBYEAR < 2025 AND (LIMIT-TO(SUBJAREA, 'ENGI')) AND (EXCLUDE(DOCTYPE, 'cr'))
Query Extensions	To focus on specific damage types and data modalities, extensions such as TITLE-ABS-KEY((crack*) AND (image* OR photo*)) were used.
Inclusion Criteria	1. Studies involving real concrete bridges; 2. Use of AI techniques in analysis or decision-making; 3. Relevance to structural condition monitoring or assessment; 4. Utilization of monitoring data, including images, signals, or sensor readings.
Exclusion Criteria	Studies focusing exclusively on simulations or laboratory-scale experiments without in situ application.

3 An Overview of Identified and Analyzed Articles

A total of 224 studies met the criteria mentioned in Table 1 and were retained for analysis. Each selected study was categorized according to the type of input data used (e.g., image, signal, point data) and the sKPI addressed (e.g., cracks, corrosion, vibration). Table 2 presents the two-dimensional counts of studies across both data type and sKPI categories. It can be found from the table that the majority of crack detection studies (95 out of 122) rely on image data, reflecting the popularity of vision-based inspection and UAV-supported surveys. In contrast, vibration monitoring is more evenly distributed across data types, with 7 image-based, 13 signal-based studies, and 5 point-based. Such a distribution can highlight the variety of approaches used to capture dynamic responses. Similarly, studies related to corrosion show a balanced use of all three data types, which can indicate that no single sensing modality dominates this category.

In order to visualize the information presented in Table 2 and to illustrate the marginal distribution of studies, Fig 1 was plotted. Fig 1-a highlights the overall proportions of studies by data type, while Fig 1-b shows the sKPI-related parameters. As shown in Fig 1-a, the majority of studies relied on image-based methods, which accounted for 64% of the reviewed studies. Fig 1-b shows that crack detection was the most frequently addressed sKPI, representing 54% of the studies.

Table 2. Number of publications that use AI to monitor concrete bridges

No. of papers		Data type			Sum
		Image/photo	Signal/wave	Data point	
sKPI-related parameters	Crack	95	19	8	122
	Spalling	17	5	0	22
	Corrosion	10	11	5	26
	Prestress loss	0	2	0	2
	Debonding	2	1	0	3
	Vibration	7	13	5	25
	Deformation	5	2	1	8
	Alkali-silica reaction (ASR)	2	0	0	2
Structural capacity		6	5	3	14
Sum		144	58	22	224

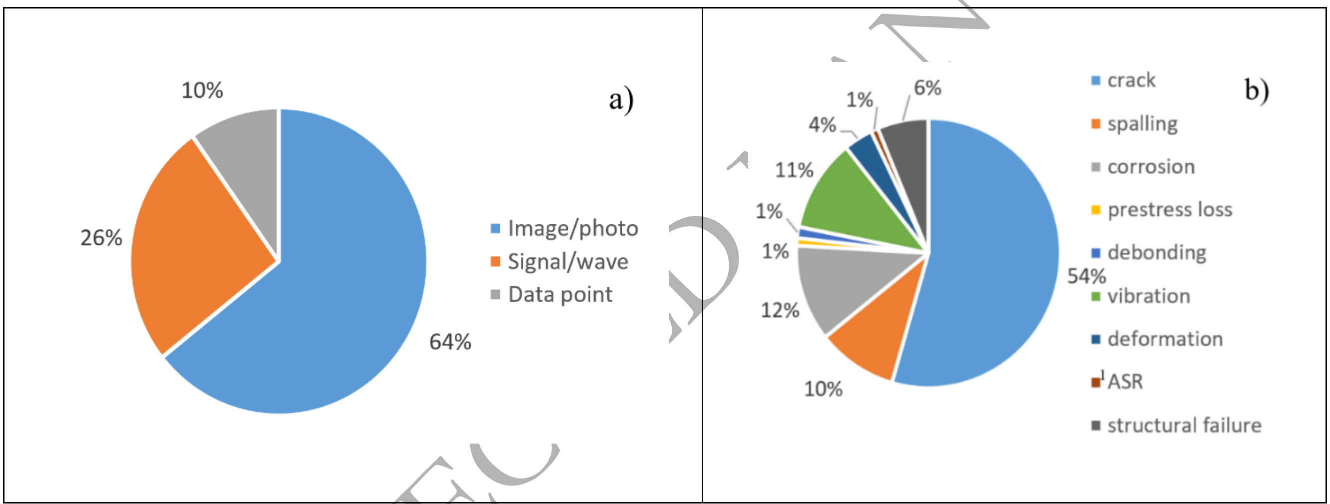

Fig 1. Distribution of publications on AI applications on (a) different data types and (b) different measured parameters in the concrete bridge health monitoring.

[1] ASR in the figure refers to the alkali-silica reaction.

In addition to the distribution of data types and sKPIs, the temporal evolution of research trends in concrete bridge health monitoring is shown in Fig 2. The corresponding figure presents the total number of studies on concrete bridge health monitoring (blue line), the subset of studies that applied AI techniques (orange line), and the annual share of AI-based studies within the total (grey bars). Between 2001 and 2010, the number of publications increased steadily, followed by noticeable fluctuations in subsequent years. Before 2010, AI-related studies represented, on average, about 14% of the total. Their limited growth before 2015 can be attributed to the immaturity of AI tools, the scarcity of datasets, computational restrictions, and the prevalence of physics-based models over early AI approaches. The observed trend between 2010 and 2015 is consistent with the observations of Adeli et al. (1989) [22] and Ko et al. (2005) [23], who emphasized the need for interdisciplinary collaboration to advance AI integration into structural health monitoring. From 2015 onward, AI applications grew rapidly, reaching a peak share of nearly 20% in 2024. This growth was facilitated by the release of open-source deep learning frameworks and the emergence of affordable sensing technologies. The slight decline after 2021 likely reflects both a shift in novelty toward AI-based methods and the disruptions to research activities caused by the COVID-19 pandemic.

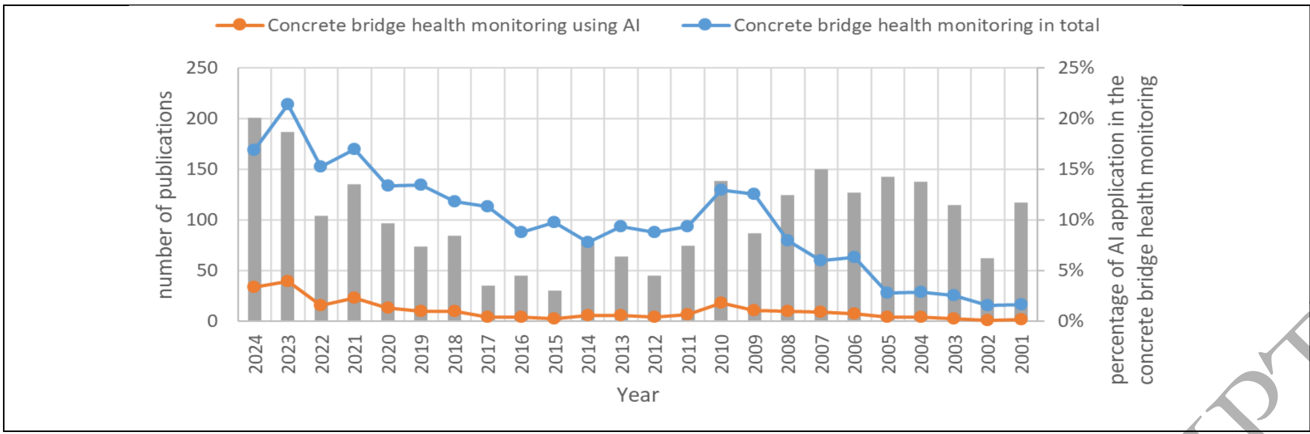

Fig 2. Annual publications on the use of AI for concrete bridge health monitoring (orange line), compared to the total number of publications on concrete bridge health monitoring (blue line). The grey bars represent the percentage of AI-related publications each year.

4 Assessment of concrete bridges based on monitoring data

This section reviews sKPIs used for the condition assessment of concrete bridges. These sKPIs differ depending on the type of bridge and the failure modes. In this review, two common types of concrete bridges, reinforced concrete slab bridges (Fig 3a) and prestressed concrete bridges (Fig 3b), were analyzed. These two types were selected for analysis due to their widespread use and distinct structural characteristics, although there are other types of concrete bridges, such as arch bridges or cable-stayed bridges [24, 25].

4.1 Structural key performance indicators for reinforced concrete slab bridges

Reinforced concrete bridges are a popular bridge structure, distinguished by their simple layout, low construction cost, and ability to carry distributed loads efficiently [26]. These bridges are often made of a reinforced concrete slab supported directly by piers, abutments, or beams, which makes them ideal for short to medium spans. This type of bridge has critical failure modes of flexural and shear failure.

Flexural failure begins when flexural cracks form at the cross-section subjected to the maximum bending moment. These cracks develop once the tensile stress at the point farthest from the neutral axis exceeds the tensile strength of the concrete. As the load increases, the crack width expands until the steel reinforcement at the crack yields. Flexural failure is triggered when the steel reinforcement yields, followed by the crushing of the concrete in the compression zone. Considering this failure mechanism, sKPIs can be defined to monitor flexural failure related to concrete and steel strain, crack location and width, and deflection [27, 28].

In flexural shear failure, as the applied load increases, flexural cracks typically initiate at the bottom of the cross-section in the shear span. These cracks usually start vertically and then incline toward the loading point, creating a flexural shear crack. If such cracks damage the strut (i.e., the compressive region between the load and the support), the structure can fail suddenly and catastrophically. Due to its brittle nature and lack of clear warning signs, flexural shear failure is challenging to predict in advance.

Fig 3. Photos of (a) a reinforced concrete slab bridge and (b) a prestressed concrete girder bridge.

Various analytical models have been developed to estimate the flexural shear capacity. These analytical models linked the flexural shear capacity with the sKPIs related to the location and width of cracks. For example, Dieteren, Bigaj-van Vliet [29] introduced a sKPI that compares the crack width in the flexural zone to the shear span of the member. Similarly, Benitez, Lantsoght and Yang [30] proposed a sKPI based on the maximum flexural crack width. In addition to crack-based indicators, some sKPIs have been developed based on parameters measured through specialized monitoring techniques such as Acoustic Emission (AE). Recent studies, for instance, have suggested a sKPI based on the number of AE events detected in the compressive zone of concrete [31].

4.2 Structural key performance indicators for prestressed concrete bridges

Due to their high strength, prestressed concrete bridges are widely implemented in the transportation infrastructure [32]. These bridges employ prestressing techniques to counteract tensile stresses brought on by external loads. In such a technique, steel tendons are tensioned to introduce compressive forces into the concrete, effectively counteracting tensile stresses caused by external loads [33]. The bending behavior of prestressed concrete bridges under the service limit state is crucial for structural safety. Most existing sKPIs are related to the strain, deformation, or vibration of the structure, assuming linear elastic behavior [7]. However, shear failure is a significant type of failure for prestressed concrete bridges, particularly when there is minimal or no shear reinforcement. A significant number of these existing prestressed bridges are considered deficient in terms of shear capacity [6, 34]. To address the issue with the shear capacity, recent research proposed the inclusion of additional sKPIs, such as the location of critical sections and stress limits, to capture potential shear-related vulnerabilities better [34].

To complement the sKPI discussions for both reinforced concrete slab bridges and prestressed concrete bridges, Table 3 presents the parameters related to sKPIs for reinforced concrete slab bridges and prestressed concrete bridges, along with their associated monitoring techniques. A detailed review of these techniques can be found in the previous study [35]. These sKPIs are mainly developed according to damage-based failure models. The damage could come from the increase in traffic load or the degradation of material properties due to corrosion or an alkali-silica reaction. Therefore, it is equally important to monitor not only the structural responses but also the underlying causes of damage, which are also listed in Table 3, to support a more comprehensive monitoring strategy.

It can be seen from Table 3 that many sKPIs across both reinforced and prestressed concrete bridges rely on similar measurement parameters, such as strain and crack width. Similar measurement parameters can suggest an opportunity for standardized sensing platforms that can be deployed across different bridge types, potentially improving cost efficiency and interoperability in SHM systems. The dependency on contact-based sensors, such as strain gauges and LVDTs, is still dominant across most failure modes. However, these sensors can be labor-intensive to install and maintain, especially in existing bridges. From the monitoring techniques column in the corresponding table, it can be observed that compared to flexural sKPIs, shear-related indicators appear less well-developed and less diverse. The observation supports the previous discussion regarding the challenge in detecting shear-related damage and highlights the need for further research into shear failure monitoring.

It should be noted that the critical locations where these sKPIs are measured are strongly dependent on the boundary conditions and the resulting stress field. In simply supported bridges, maximum flexural and shear stresses typically concentrate at midspan and near supports [36], respectively. In continuous bridges, regions near internal supports often govern due to negative bending moments and potential stress concentrations [37]. For arch bridges, the line of thrust varies under load, requiring careful monitoring of both the crown and haunches [38].

Table 3. sKPI-related parameters and the associated monitoring techniques

Structural type	Structural performance	sKPIs	Monitoring techniques
Reinforced concrete bridges	Flexural failure	Concrete and steel strain	Strain gauges, LVDT, fibre optical sensors (FOS)
		Crack location and width	Crack gauges, Digital Image Correlation (DIC), photogrammetry, LVDT, Extensometers, FOS, acoustic emission (AE)
		Deflection	Satellite, LVDT, Geodesy, DIC, Laser distance finder, FOS
	Flexural shear failure	Crack location and width	Crack gauges, Digital Image Correlation (DIC), photogrammetry, LVDT, Extensometers, FOS, acoustic emission (AE)
Prestressed concrete bridges	Flexural behaviour	Concrete strain	Strain gauges, LVDT, FOS
		Vibration	Accelerometers
		Deflection	Satellite, LVDT, Geodesy, DIC, Laser distance finder, FOS
	Shear failure	Stress and its location	Strain gauges, LVDT, FOS
Others	--	Alkali silica reaction (location, amount)	Crack gauges, Electrical impedance
		Corrosion (location, area)	Half-cell potential, Concrete resistivity, Infrared thermography, AE, X-ray
		Load	Weigh-in-motion (WIM)

5 General overview of AI approaches for interpreting monitoring data

The raw data acquired through the monitoring techniques listed in Table 3 has various types such as images, signals, and point data. Several essential steps are required to transform this diverse data into meaningful insights that support the assessment of structural performance and the derivation of sKPIs. These steps include data pre-processing, feature extraction, and data processing for clustering, classification, regression, and optimization, followed by potential post-processing to improve visualization. While these steps are not exclusive to AI-based workflows, they are increasingly supported by AI tools that will enhance automation, pattern recognition, and predictive modelling.

Fig 4 shows the general AI implementation steps within the context of SHM and the techniques commonly applied at each stage. The stages and techniques represented in the figure reflect their established use in bridge SHM and similar engineering applications, where data quality issues, high dimensionality, and the need for interpretable results are common challenges. This section provides a brief introduction to the fundamentals of these AI implementation steps, serving as a reference for the subsequent sections.

5.1 Data pre-processing

Data pre-processing involves preparing raw data for further analysis by cleaning, transforming, and organizing it to ensure the accuracy, consistency, and usability of the data. Clean and well-processed data can lead to more accurate and robust AI models. Moreover, high-quality data provides better insights, leading to more informed decisions. The key aspects of data pre-processing are briefly presented in the following subsections.

5.1.1 Data Cleaning

Raw data often has missing values, an imbalanced distribution, and various types of errors [39]. To ensure high-quality input for analysis, data cleaning and augmentation techniques are often applied [40]. Data cleaning involves handling missing values, removing outliers, and resolving discrepancies in the dataset [41]. Data augmentation creates new data samples through operations, such as rotation, flipping, or scaling [42]. Additionally, data augmentation can help balance underrepresented groups within a dataset, which is essential for reducing bias and ensuring accurate model training [40, 43].

5.1.2 Data Transformation

Data transformation involves converting data into a format or structure that is suitable for analysis. A common step in data transformation is scaling, which includes techniques such as normalization and standardization. Variables within a dataset often have different scales. For example, bridge displacement might range from 0 to 10 mm, while temperature can range from 10 to 30 degrees Celsius. Without proper scaling, AI algorithms may incorrectly assign greater importance (weight) to variables with larger numerical ranges. Normalization and standardization can address this by bringing variables to the same scale [44].

Normalization scales all values to a range between 0 and 1, whereas standardization transforms the data to have a mean of 0 and a standard deviation of 1 [44]. Choosing between normalization and standardization depends on the characteristics of the data and the implemented algorithm. Some AI models, such as tree-based algorithms, generally tend not to be impacted by feature scaling and do not necessitate these transformations [45].

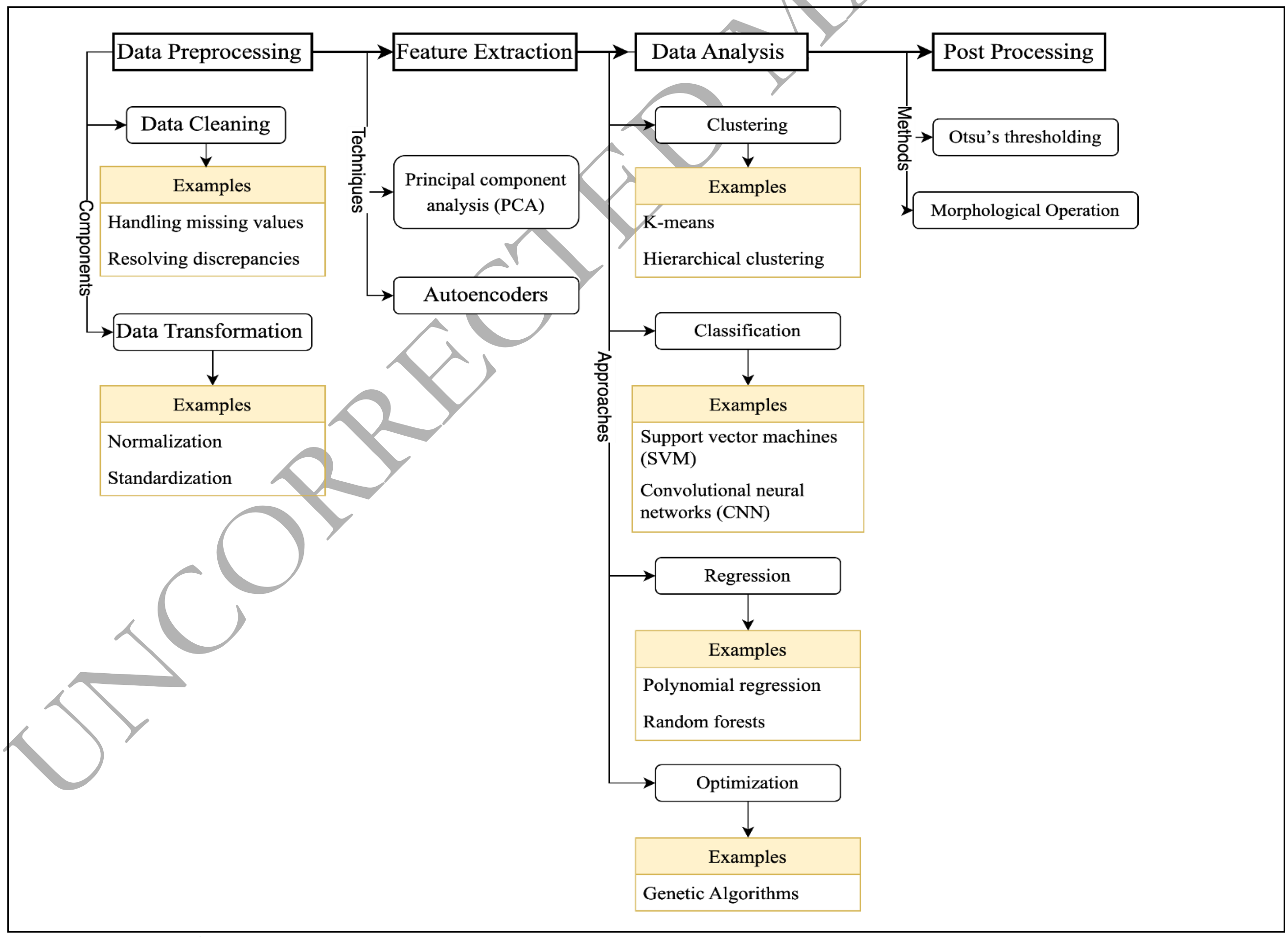

Fig 4. Steps to use AI for interpreting monitoring data.

Besides normalization and standardization, another aspect of data transformation is adapting the data format to meet the specifications of particular AI models. In particular, AI algorithms such as convolutional neural networks (CNNs) are often more effective when working with two-dimensional (2D) data such as images. As a result, one-dimensional (1D) data, such as time-series signals or point measurements, is frequently transformed into a 2D format to leverage these models.

The appropriate transformation technique depends on the nature of the 1D data. For instance, signals that carry frequency information, such as AE data, can be converted into spectrograms [46]. A spectrogram is a 2D visual representation in which time is shown on the *x*-axis, frequency on the *y*-axis, and color intensity indicates the amplitude of specific frequencies over time. For other types of 1D data, such as fixed-frequency signals where frequency content is limited or not of interest, simpler transformations like reshaping, windowing, or segmentation may be applied. It is noted that these transformations do not add new information to the data; rather, they only restructure the data to make it more compatible with certain AI algorithms.

In bridge health monitoring, the selection of preprocessing techniques is often driven by the type of data and the practical challenges it presents. For image data, noise often arises from issues such as uneven lighting or shadows, making detection challenging. To address these issues, researchers have applied techniques such as sharpening and smoothing, illumination correction, segmentation methods and shadow-removal. Although these techniques enhance damage visibility and reduce noise effectively in controlled laboratory settings, their robustness often remains limited when applied to field-acquired images.[47, 48].

Similar challenges in image data often occur in time-series data, where noise can obscure the structural response of a bridge. To alleviate obscurity in structural responses, noise filtering is considered a critical preprocessing technique. Ye and Chen (2018) [49] employed a Butterworth low-pass filter technique to separate high-frequency live load effects caused by vehicles from the low-frequency structural deflections. This technique not only clarified the signal but also improved the stability of subsequent feature extraction techniques. Strain measurements face comparable issues, as they are often contaminated with noisy data points. Akintunde et al. (2021) [50] addressed this by centering the data to remove the mean and whitening it to normalize the variance before applying Independent Component Analysis (ICA) for feature extraction.

Beyond noise, incomplete data presents another major challenge in bridge health monitoring. One strategy is data augmentation, which artificially expands the dataset while mimicking real-world variability [51, 52]. For image data, common augmentation techniques include random-resized cropping, random rotation, colour jittering, and horizontal flipping. These methods increase the diversity of the dataset and reduce the risk of overfitting.

5.2 Feature extraction

Following the preprocessing, the subsequent step in the AI pipeline for SHM is feature extraction. In feature extraction, the aim is to obtain more valuable or representative information from the source data for downstream learning tasks. The goal is to represent the data more effectively by transforming complex data into essential features well-suited for AI algorithms. By using the essential features, the dimensionality of the dataset can be reduced, which in turn improves model efficiency, interpretability, and overall performance. Although a wide variety of feature extraction techniques are available, selecting an appropriate technique depends on the data type and the specific requirements of the subsequent learning task.

In bridge health monitoring, one commonly applied method for feature extraction is Principal Component Analysis (PCA) [53]. It is a statistical technique to decrease the dimensionality of a dataset by projecting it onto a new coordinate system. In PCA, the first principal component corresponds to the direction of the highest variance in the dataset. Each successive principal component forms a new axis that is orthogonal to all previous ones and captures the next highest variance. Variants such as kernel PCA expand the original PCA approach to capture nonlinear structures in the dataset [54]. An application of PCA [47] combined handcrafted features with kernel PCA for concrete crack classification. Zoubir et al. (2022) [47] combined handcrafted features with kernel PCA for concrete crack classification. Parameters were reduced from thousands of dimensions to 100 components using PCA with an RBF kernel. The value of nonlinear dimensionality reduction was shown by improved accuracy (up to 99.26%) in crack detection.

Another widely used feature extraction method in bridge health monitoring is the autoencoder, a neural network designed to learn compressed representations of input data [55]. Autoencoders consist of an encoder, which encodes the input data into a compressed representation, and a decoder, which reconstructs the input data from the compressed representation. Researchers in [56] used an encoder of 13 layers for feature extraction from concrete crack images. The encoder was able to extract hierarchical features, including low-level features (e.g., edges, corners, textures) in earlier stages and high-level abstract features (i.e., patterns and shapes of cracks) in deeper layers.

5.3 Data analysis based on features

Following the preparation of the monitored dataset, AI techniques can be applied to analyze and interpret the data. These techniques can be categorized as clustering, classification, regression, and optimization, each serving different analytical purposes, such as pattern recognition, prediction, decision-making, and model tuning. The choice of techniques depends on the nature of the data, the task purpose, and the desired outcome of the analysis. In the following subsections, a brief overview of each of these AI techniques is provided.

5.3.1 Clustering

Clustering is an unsupervised learning method that groups the data based on their similarities without prior labels, aiding in discovering underlying patterns and segmenting datasets for further analysis [57]. Some commonly used clustering methods include *K*-means [58] and hierarchical clustering [59]. *K*-means partitions "*n*" data points into "*k*" clusters, assigning each point to the nearest centroid. The algorithm starts with initial centroids, iteratively reassigns points to the closest centroid, and updates centroids until stabilization or a set number of iterations. A key drawback in *K*-means clustering is the need to pre-specify "*k*" and sensitivity to initial centroids [58]. Hierarchical clustering creates a cluster hierarchy using either a bottom-up (agglomerative) or top-down (divisive) approach. Agglomerative clustering starts with each point as its cluster, merging pairs upward [59], while divisive clustering starts with one cluster, splitting downward [60]. The process is visualized with a dendrogram, a tree-like diagram representing the clustering hierarchy and providing a visual summary.

5.3.2 Classification

Classification is a supervised learning method that assigns data to predefined classes or labels [61]. Common classification techniques include support vector machines (SVM) [62], ensemble learning, and convolutional neural networks (CNN) [63]. SVM classifies data by finding an optimal hyperplane in an N-dimensional space, where "N" is the number of features, that maximizes the margin between different classes. Each data point is represented as a coordinate in this space based on its feature values [62].

Ensemble learning is a method whereby several models, often referred to as 'weak learners', are aggregated to create a robust predictive model. In ensemble learning, the basic concept is that the final model can attain superior prediction performance and robustness than any individual model (weak learners) by aggregating their predictions. Ensemble learning is particularly useful in reducing variance and bias and improving the generalization of the model. Common examples of ensemble learning methods include bagging, boosting, and stacking. Algorithms such as the Decision tree [64] and random forest [65] are commonly used as weak learners in ensemble frameworks.

CNNs are a type of deep learning model frequently used for processing grid-like data, such as images [63]. They are designed to capture spatial patterns in data and typically operate on inputs with three dimensions: width and height (representing the spatial layout), and depth (representing the number of channels, such as color components in an image). CNNs have shown remarkable success on problems involving image and video recognition. Similar to traditional multilayer neural networks, a standard CNN architecture consists of one or more convolutional layers, often coupled with subsampling (or pooling) layers, and then one or more fully connected layers. The convolutional layers enable CNNs to automatically learn spatial hierarchies of features, making them especially effective for image-based SHM applications.

The CNNs can be enhanced through transfer learning, which can reduce computational costs and improve training efficiency [66]. Instead of training a model from scratch, in the transfer learning approach, parameters from the convolutional layers of a pre-trained neural network are reused. [14]. Commonly used pre-trained models include VGG16, ResNet18, DenseNet, AlexNet, MobileNet, Inception v3, etc. [67].

In image classification tasks, instead of classifying the entire image, another strategy is to classify individual pixels, which is called semantic segmentation. Semantic segmentation provides a pixel-level classification, which helps to understand the image at a much more granular level. CNNs are commonly used for semantic segmentation due to their effectiveness in dealing with image data. In particular, fully convolutional networks (FCN) are designed for semantic segmentation [68]. The FCN takes an input image and produces an output image of the same size, where each pixel is classified into a certain class. More advanced techniques like U-Net, SegNet, and DeepLab have also been proposed to improve the performance of semantic segmentation.

5.3.3 Regression

While classification is used to classify categorical values, regression is designed to predict continuous numerical values [69]. Similar to classification, regression is a supervised learning method, as it aims to model the relationship between a dependent variable (also called the target or outcome) and one or more independent variables (also called features or predictors). Traditional regression methods encompass linear regression, polynomial regression, and logistic regression, although the latter is often used for classification purposes. As both regression and classification fall under supervised learning, many AI methods used for classification, such as SVM, decision trees, random forests, and neural networks, are equally applicable to regression.

5.3.4 Optimization

Optimization is a fundamental concept in AI and ML that focuses on finding the best set of parameters that minimizes or maximizes a given objective function [70]. One widely used AI-based optimization technique is the genetic algorithm (GA) [71]. GAs are adaptive heuristic search algorithms inspired by the principles of genetics and Darwinian natural selection. They solve optimization problems by evolving a population of candidate solutions over successive generations, using biologically inspired operators such as selection, crossover (recombination), mutation, and inheritance. GAs are particularly useful for solving complex, nonlinear, and multi-objective optimization problems where traditional methods may be inefficient or inapplicable.

The suitability of the techniques discussed in Section 5.3 for SHM tasks depends on factors such as the amount of training data, the availability of computational resources, and the specific sKPIs being assessed. In the reviewed literature, CNNs were the most commonly applied method for crack identification. Approximately 30 percent of the studies employed standard CNN architectures, while 20 percent made use of FCNs. A further 20 percent applied transfer learning or lightweight CNN variants to improve computational efficiency, and the remaining studies relied on other techniques such as SVMs. For estimating deflections and structural capacity, the predominant methods included SVM, regression models, and ensemble learning algorithms. To enhance model performance, several studies implemented GAs to optimize the parameters of AI models for capacity prediction.

Table 4 summarizes the typical use cases of the techniques mentioned above for different types of SHM data, along with their strengths and limitations. A more detailed evaluation regarding the performance of these AI techniques to detect different sKPIs is presented in Section 6.

5.4 Post-processing of the results

Postprocessing techniques are frequently applied in bridge health monitoring to improve the visualization and interpretability of results, particularly by refining crack boundaries. Unlike preprocessing or detection methods, these procedures typically do not rely on AI but instead use established image-processing approaches. One widely used method is Otsu's thresholding, which automatically separates image regions based on pixel intensity [72]. For example, Li et al.

(2018) [73] applied a fixed threshold of 0.5 to the pixel-level confidence map generated by a CNN model, classifying each pixel as either crack or non-crack. The binarized crack map was then refined through noise removal, discarding isolated pixel clusters and producing a thinner, more continuous crack path that closely matched the actual defect.

Morphological operations represent another category of postprocessing techniques, offering tools such as colour correction, median filtering, line enhancement, and binarization to reduce noise, enhance contrast, and emphasize crack-like features. For example, Nomura et al. (2022) [74] improved YOLOv2 crack detections by applying a sequence of morphological operations to achieve pixel-level crack mapping. Colour correction and filtering reduced background noise, Hessian-based line enhancement emphasized linear structures, and Canny binarization produced cleaner crack masks. These steps transformed coarse bounding boxes into detailed crack geometries, thereby improving the accuracy and reliability of crack characterization.

Table 4. Evaluating the performance of AI techniques for different types of SHM data

SHM Data Type / Task	Preferred Model	Strengths	Considerations
Crack images (RGB / thermal)	CNN / FCN / U-Net	Capable of learning crack location and morphology directly from raw image pixels; robust to lighting variations with appropriate tuning	Requires annotated datasets and sufficient computational resources (e.g., GPUs) for effective training
Vibration time-series	SVM	SVM works efficiently in low-data regimes	SVM performance may degrade when the size of the datasets increases significantly
Mixed sensors (strain, temperature, modal data)	Random Forest / Gradient Boosting	Handles nonlinear feature interactions; offers feature importance insights for interpretation	May benefit from prior feature selection or dimensionality reduction if features are highly correlated or redundant
Parameter tuning/model optimization	Genetic Algorithm	Enables exploration of complex, high-dimensional parameter spaces without requiring gradient information	Convergence rate and performance may be sensitive to the choice of objective function and the scale of the problem

6 Evaluation of AI performance for concrete bridge assessment

Building on the previous discussion of AI techniques and data processing, this section evaluates the performance of AI in real-world concrete bridge health monitoring and assessment. The evaluation focuses on key criteria such as accuracy, efficiency, and robustness, as well as the use of AI for structural capacity estimation and advanced applications. In the following subsections, a brief explanation of each aspect is provided, highlighting recent developments, example applications, and remaining challenges.

6.1 Using AI to identify damage in concrete bridges

A large proportion of the studies reviewed in this paper used AI techniques to process image data to detect damage like cracks. Accordingly, the primary focus in this subsection is to evaluate the effectiveness of AI in identifying concrete damages, predominantly using image data. An effective method for damage identification should offer sufficient accuracy, support real-time monitoring, and remain robust under complex environmental conditions. Thus, the evaluation focuses on these characteristics under varying environments, advantages over non-AI methods, and limitations.

6.1.1 Accuracy of the damage detection

The performance of crack detection models is commonly evaluated using standard classification metrics such as accuracy, precision, recall, and F1-score. In some cases, variant or derived metrics may also be employed, depending on the specific application or dataset. Accuracy measures the proportion of correctly identified instances among all predictions. Precision quantifies the proportion of true positive detections among all instances classified as positive by the model, indicating the number of correctly detected cracks. Recall (also known as detectability measures the proportion of actual cracks that were correctly detected by the model. The F1-score is the harmonic mean of precision and recall, and provides a balanced assessment, especially in imbalanced datasets where one class may dominate [75].

In a study by Zoubir et al.(2022) [76], the performance of SVM, decision tree, and random forest classification in the crack detection was evaluated. The dataset used in the study consisted of images taken from an actual bridge inspection. Image pre-processing, feature extraction, and dimensionality reduction were first performed on the collected data. Then, the classification performance was evaluated, reporting that SVM and random forest performed better than the decision tree in accuracy, precision, recall, and F1. Furthermore, it was reported that the dimensionality reduction approach enhanced classification performance by mitigating the curse of dimensionality, reducing noise, and isolating the most relevant features.

Modarres et al. (2018) [77] compared the performance of CNN, SVM, and random forest in the crack detection, using two different scenarios. The first scenario was the classification of clean images of well-controlled concrete surfaces, where some light staining (i.e., minor discoloration or surface contamination such as dirt or watermarks) was present. The second scenario was the classification of noisy images of weathered and highly textured concrete surfaces, where staining was more predominant. The second scenario represented the situation of a field inspection. The authors reported that in both scenarios, CNN showed higher accuracy than the other methods.

Pozzer et al. (2021) [78] compared the performance of five CNN models for the detection of multiple types of damages, including cracks, delamination, spalling, and patches. The five CNN models were VGG 16, ResNet 18, ResNet 50, Xception, and MobileNetV2. The results showed that each CNN model has better performance in one of the accuracy metrics. For example, MobileNetV2 performed the best in the recall metric, while VGG 16 had the highest performance in precision. Therefore, the authors emphasized that the selection of the AI models depends on the objective of the study.

Kim et.al. (2019) used a mask and a region-based convolutional neural network (Mast R-CNN) to detect concrete cracks [51]. The trained model can successfully find cracks with widths of 0.3 mm or more with errors less than 0.1 mm. The cracks with a width less than 0.3 mm show relatively larger errors due to the limitation of image resolution.

Many researchers applied semantic segmentation to detect finer cracks, which classifies the damage at the pixel level. For example, Manjurul et al.(2019) [56] used a FCN with an encoder and decoder framework for semantic segmentation. The FCN is trained using an open-source concrete crack image dataset. Benchmarking with the results of SVM and CNN using the same dataset, the FCN showed higher precision, recall, and F1. Liu et al.(2019) [79] used U-Net, an improvement of FCN, to detect cracks. Compared to other CNN models, U-Net made a pixel-level detection with higher resolution. And U-net showed robust performance under dark conditions and a noisy background. Furthermore, compared to other FCN models, U-Net reached higher precision, recall, and F1 with a smaller training set.

Adel et al. [80] (2021) utilized U-Net to detect pits in images of concrete cracks. They were able to recognize pits ranging from 0.15 mm to 8 mm in size. The density of these pits was found to be related to the shear deformation of the concrete, thus implying a relationship to the overall structural performance. Near failure, an increase in pit density could be an early warning sign. However, the verification was carried out in controlled laboratory conditions. In real-world scenarios, factors such as surface dirt on the structure could introduce background noise that interferes with pit detection. Therefore, the robustness of this method to such noise still needs to be improved.

6.1.2 Efficiency

One of the major limitations of AI applications in damage detection is their high computational cost, particularly in deep learning-based methods [81]. To address the trade-off between model complexity and efficiency, several studies have explored architectural optimizations. Lopez Droguett et al. (2023) [82] investigated how reducing the number of parameters in semantic segmentation models affects performance. They proposed two strategies: replacing the classification component with a lightweight CNN (e.g., LeNet-5) and simplifying DenseNet by limiting the architecture to just 13 layers. Their findings showed that segmentation accuracy might be maintained while still obtaining notable simplification of the model complexity. Similarly, Ren et al. (2017) [83] improved the efficiency of object detection by introducing Faster R-CNN. This model replaced the computationally expensive selective search algorithm with a Region Proposal Network (RPN), enabling faster and more accurate region detection. Following this, several real-time object detection frameworks, such as YOLO, SSD, and RetinaNet, were developed to further improve the balance between speed and accuracy.

To support real-world deployment, efforts have also been made to adapt CNNs for low-resource environments. Zhang et al. (2022) [84] restructured a CNN with millions of parameters to run efficiently on an edge device. They removed redundant layers and replaced them with smaller, computationally lighter ones, resulting in a model with fewer parameters and faster processing times while maintaining acceptable performance levels. Transfer learning has also emerged as a promising approach to enhance computational efficiency. Miao et al. (2021) [14] developed a damage detection framework called Damage-Net, combining VGG-16 and U-Net architectures. By applying transfer learning instead of training the network from scratch, they reduced computing time and memory usage by approximately 25%, without sacrificing accuracy. Finally, the application of AI for automated crack detection has demonstrated not only accuracy but also time savings in field scenarios. Lee et al. (2021) [85] employed a CNN-based method for on-site concrete crack detection using a mobile device. Images were captured in the field and analyzed on a web server. Compared to manual inspection, which took 17 minutes, the automated method completed the same task in only 4 minutes, highlighting its potential for rapid, on-site evaluation.

6.1.3 Robustness in the real environment in the field

A key challenge in applying AI models for damage detection is ensuring their ability to generalize beyond controlled laboratory settings to complex, dynamic real-world environments [86]. Models trained in ideal conditions, such as images captured under uniform lighting and clean backgrounds, often struggle when deployed in the field, where lighting, weather conditions, surface textures, and background noise can vary significantly. Crack detection, in particular, relies heavily on image contrast to distinguish cracks from the surrounding surface.

AI models trained exclusively on clear, laboratory-generated data, which lack environmental noise, often fail to detect cracks accurately during on-site inspections. Pal et al. (2021) [87] demonstrated that shadows can lead to false positives or false negatives in CNNs, significantly reducing detection accuracy, from nearly 100% to around 50%. The study also noted that standard preprocessing techniques, such as shadow removal, may degrade image quality and obscure crack features, further reducing model performance. Therefore, several studies have demonstrated that training AI models with datasets incorporating real-world noise, such as varied illumination, stains, and motion blur, can significantly improve their accuracy in the field. Modarres et al. (2018) [77] compared CNN performance on two datasets: a 'clean' set containing clear images and a 'noisy' set designed to simulate field conditions with weathered, textured concrete and significant staining. The CNN achieved 99.6% accuracy on the clean set and 98.8% on the noisy set, outperforming traditional machine learning algorithms such as logistic regression, SVM, random forest, and multilayer perceptron (MLP). To further test robustness, the authors trained the CNN on mixed datasets of clean and noisy images and evaluated it only on noisy images. The CNN maintained high accuracy (up to 97.9%), precision, and recall, consistently outperforming the MLP.

Building on the importance of robustness under noisy conditions, several innovative AI approaches have been developed. Li et al. (2017) [88] proposed an SVM combined with a greedy search strategy that filtered noise while preserving crack edges. Tested on 1,200 images from 10 real bridges, their method outperformed contemporary algorithms, achieving an average F1-score of 0.86 with a processing time of only 128 ms per image. The approach also measured crack width with an error of less than 0.03 mm, even under environmental noise. More recently, Li et al. (2020) [89] introduced an NB-FCN model that fuses a fully convolutional network with naïve Bayes data fusion to detect cracks in the presence of handwriting, water stains, surface peeling, and repair traces. To improve robustness, their framework included extensive data augmentation, such as image rotation and Gaussian-based brightness variations. When evaluated on 7,200 images, the NB-FCN achieved an average error rate of 1.28%, outperforming both CNNs and standard FCNs.

6.2 Using AI to estimate the structural capacity

In addition to identifying localized damage as discussed in the previous sections, recent AI applications aim to estimate the structural capacity to resist applied loads without failure [90]. Structural capacity provides a holistic measure of structural integrity, while traditional indicators, such as cracks, spalling, or deflection, often capture only isolated aspects of deterioration [91]. In many cases, AI models are being explored as alternatives to physical structural models due to their lower computational demand [92], greater adaptability to dynamic conditions [93], and potential for integration with historical data to enhance prediction accuracy.

Kumar et al. (2023) [94] used several ML models, including ANN, Adaptive Neuro-Fuzzy Inference Systems (ANFIS), decision trees, and XGBoost, to predict the shear capacity of reinforced concrete (RC) beams. These models outperformed traditional analytical approaches, and key influencing parameters identified included the effective depth of the beam, stirrup spacing, and shear span-to-depth ratio.

Fathalla et al. (2018) [95] trained an ANN using data from multiscale simulations to relate crack patterns and fatigue progression to the remaining life of concrete slabs. Although the model successfully reduced reliance on computationally expensive simulations, its applicability was limited by its failure to consider variations in slab geometry, reinforcement, and material properties. Additionally, the model could not differentiate among failure modes, relying solely on surface-level crack patterns.

Similarly, Lattanzi et al. (2016) [96] used regression techniques to correlate visual damage in 2D images with peak seismic displacement in columns. While strong correlations were observed, the generalizability of the model to different structural types or damage mechanisms was not validated. The complexity of failure modes highlighted the insufficiency of image data alone for robust structural assessment.

Expanding on this idea, Davoudi et al. (2018) [97] used image-based regression models, Gaussian process regression, SVM, and ANN, to estimate internal force levels (e.g., shear and moment). They found that geometric parameters, such as span-to-depth ratio, significantly improved model accuracy, underscoring the importance of contextual information in predictive performance due to scale effects in concrete behavior.

To further enhance prediction accuracy, Jiang et al. (2024) [98] applied a hybrid approach by combining SVM with GA for hyperparameter tuning. Their SVM-GA model outperformed other models, including random forests and ANN. Zhou et al. (2024) [99] validated similar models using real data from two in-service concrete bridges, demonstrating the potential for field application.

Beyond fully replacing physical models, AI has also been applied to support model updating. For example, Locke et al. [100] integrated AI into the finite element modeling (FEM) process using Bayesian estimation to calibrate uncertain parameters such as material properties and boundary conditions. This hybrid approach allowed the FEM to better reflect actual structural behavior based on monitoring data.

6.3 Broader capabilities of AI in bridge assessment

In addition to estimating structural capacity and detecting localized damage, AI offers several other capabilities that enhance bridge assessment. One notable advantage is its ability to distinguish among multiple types of damage, such as cracking, spalling, rebar exposure, and buckling. For example, Miao et al. (2021) [14] developed two models, Crack-Net and FourCategory-Net, to detect cracks and four additional damage types: spalling, rebar exposure, buckling, and fracture. Both models demonstrated strong performance, achieving a mean Intersection-over-Union (IoU) of over 70%.

AI also supports signal decomposition and interpretation for complex structural behaviors. Ye et al. (2018) [49] applied AI to separate deflection data influenced by structural damage, temperature fluctuations, and traffic loading, an essential step for accurate structural assessment. Initially, high-frequency noise from traffic loads was filtered. Then, Ensemble Empirical Mode Decomposition (EEMD) was used to extract intrinsic mode functions (IMFs), and PCA was applied to reduce feature dimensionality. Finally, Fast Independent Component Analysis (FastICA) was used to isolate independent deflection components. This method was validated through simulations and real-world tests, showing reliable performance for noise levels under 10%. However, real-world application accuracy may be affected by sensor faults, which require correction.

AI also contributes to the development of new sKPIs for condition assessment. Akintunde et al. (2021) [50] applied unsupervised ML to derive novel sKPIs from strain measurements on concrete bridge mock-ups. Using feature extraction techniques like singular value decomposition (SVD) and independent component analysis (ICA), the authors developed indices capable of distinguishing between uncracked and various damaged conditions. While the results were promising, broader application remains limited due to the statistical nature of SVD and the need to validate the method across more damage types.

Another important application of AI is data integration from multiple non-destructive testing (NDT) methods. Anterrieu et al. (2019) [101] used a decision tree learning approach to integrate diverse measurements for corrosion rate estimation, including ground penetrating radar, electrical capacity, and half-cell potential. The decision tree model performed comparably to traditional data fusion techniques such as cokriging [102], Bayesian Sequential Simulation (BSS), and classical NDT fusion methods. Its main advantage lies in its ability to utilize all available data in a unified framework. However, it requires a comprehensive database for training.

To complement the studies reviewed in this section, two summary tables (see Table 5 and Table 6) compare the main AI methods across different data types. Table 5 outlines AI applications for crack identification from images, while Table 6 focuses on structural capacity estimation and hybrid updating. Each table highlights the respective strengths, limitations, accuracy trends, and computational requirements, providing a concise overview of the trade-offs involved in selecting methods for bridge condition assessment. While these tables synthesize the performance of individual AI techniques in isolated tasks, real-world bridge assessment often requires more comprehensive solutions. Accordingly, the next section shifts the focus from discrete methods to integrated, end-to-end frameworks that combine multiple techniques to address the multifaceted challenges of bridge assessment.

Table 5. Performance of AI techniques for crack identification (sKPI).

Method	Strengths	Limitations	Accuracy (vs. peers)	Computational profile	References
SVM / Decision Tree / RF	Simple and interpretable; effective with small datasets and after dimensionality reduction	Performance depends strongly on feature design; limited robustness under noise and shadows	SVM / RF ≥ DT; lower than CNNs in complex/noisy scenes	Lightweight; CPU-friendly; fast training and inference	[47, 48, 88, 103-105]
CNN classifiers (VGG, ResNet)	High accuracy and robustness; transfer learning improves efficiency	Computationally intensive; less effective for detecting fine cracks	Consistently > SVM/RF on clean and noisy images	Moderate–high; GPU preferred; transfer learning reduces demand	[73, 77, 78, 106-108]
Lightweight CNNs (e.g., MobileNet)	High recall with fewer parameters; suitable for edge and mobile applications	Slight reduction in precision under challenging conditions; limited for detecting subtle or complex crack patterns	High recall; precision slightly below VGG/ResNet	Lightweight; deployable on edge/mobile devices	[82, 84]
Two-stage detectors (Faster/Mask R-CNN)	Accurate region proposals; Mask R-CNN capable of estimating crack widths above 0.3 mm	Computationally expensive; limited resolution for detecting hairline cracks	Strong localisation; sub-mm estimates above 0.3 mm	Heavy; GPU required; slower than one-stage	[51, 83]
One-stage detectors (YOLO, SSD, RetinaNet)	Enable real-time detection with a good balance between speed and accuracy	Less effective for detecting fine cracks	Enables rapid field triage with acceptable accuracy	Moderate; real-time feasible on GPUs/edge devices	[74, 85]
Semantic segmentation (FCN, U-Net, DeepLab)	Provide pixel-level mapping; effective for thin cracks and area quantification; U-Net performs well with relatively small datasets	Require large amounts of annotated data; performance declines under severe shadows or surface soiling	U-Net/FCN > SVM/CNN baselines; robust in noisy/dark scenes	Moderate–high; GPU preferred; feasible with transfer learning	[56, 79, 80, 89, 109]

Table 6. Performance of AI techniques for structural capacity estimation (sKPI).

Method family	Strengths	Limitations	Accuracy (vs. peers)	Computational profile	References
Linear / Elastic-net regression	Transparent; fast, low data demand	Cannot capture nonlinear effects or interactions	Baseline; typically outperformed by nonlinear ML	Very lightweight; CPU-friendly	[96]
SVM (regression/classification)	Performs well with small datasets; good margin separation	Sensitive to kernel choice and scaling	Competitive; outperformed trees in some cases; improved with GA tuning	Moderate; CPU/GPU; training cost grows with data size	[97, 110]
Decision Trees / RF / XGBoost	Handles heterogeneous features; provides feature importance	Can overfit; weaker extrapolation	Often top-performing classical baselines for shear/capacity	Moderate; CPU-friendly; fast inference	[94, 110]
ANN / MLP	Captures nonlinearities and feature interactions	Requires more data; tuning sensitive; less interpretable	Outperforms analytical formulas; accuracy depends on data completeness	Moderate; GPU useful for larger networks	[94, 95, 110]
Gaussian Process Regression (GPR)	Provides uncertainty estimates; strong with small data	Cubic[1] scaling; impractical for large datasets	Competitive with informative features; robust on limited data	Heavy for large N; CPU/GPU	[97]
Hybrid SVM–GA / meta-heuristics	Optimised hyperparameters; better margins	Extra tuning layer; reproducibility issues	Reported gains over RF/ANN on the same datasets	Moderate; added tuning cost	[98, 99]
FEM + Bayesian updating (hybrid)	Physics-consistent; interpretable; integrates prior knowledge	Requires detailed modelling and prior assumptions	Improves fidelity to real behaviour; complements data-driven models	Moderate–heavy for calibration; light inference once trained	[100, 110]

[1] GPR scales cubically with dataset size ($O(n^3)$), which limits its use to small datasets.

6.4 Advanced System-Level AI Frameworks for Bridge Assessment

Building on the previous review, this section focuses on innovative AI frameworks applied in concrete bridge health monitoring. A key insight from the literature is that many advanced implementations involve integrating multiple AI algorithms and techniques into cohesive frameworks, enhancing detection accuracy and system robustness.

Prateek et al. (2016) [103] proposed a noise-resilient AI framework for crack detection. The framework began with a robust line segment detector to accurately extract crack patterns, followed by feature selection and classification using an SVM. This combination improved detection performance under noisy conditions.

Nomura et al. (2022) [74] developed a hybrid system by integrating the YOLO (You Only Look Once) object detection algorithm with VGG16, a CNN-based image classifier, along with several image processing techniques. YOLO was used to rapidly identify candidate regions in images, enabling real-time monitoring. These regions were then refined using VGG16 for pixel-level crack recognition. Additional image enhancement methods, including color tone correction, median filtering, line enhancement, and binarization, were applied to improve detection clarity.

Ni et al. (2019) [107] proposed a dual-scale CNN architecture by combining GoogLeNet and ResNet20 for crack detection. GoogLeNet was used to scan large images and extract small windows containing potential cracks. These windows were then passed to ResNet20 for finer-scale analysis to identify thinner cracks. While this dual-scale approach improved detection sensitivity, it was more computationally demanding compared to single-scale models.

Ma et al. (2022) [109] proposed an 'artificial-and-real' parallel system to enhance bridge inspection processes. In this system, the AI model was continuously updated with incoming data, and its predictions informed the planning of subsequent data acquisition activities. The AI-based and real (physical inspection) systems co-evolved in parallel, improving both the efficiency and accuracy of inspections.

To complement the preceding evaluation, selected case studies are presented to demonstrate the practical application of AI in bridge health monitoring. These examples provide concrete illustrations of the implementation process, highlight the outcomes achieved, and expose challenges encountered under real-world conditions.

6.5 Illustrative case studies

In this section, three representative applications of bridge structural health monitoring are considered: (i) damage identification using image-based crack segmentation, (ii) capacity prediction using GA-enhanced SVM, and (iii) portable computer–vision–based load rating. Each case study is discussed in the following subsections, and a summary is provided in Table 7.

6.5.1 Case study i: Damage identification using image-based crack segmentation

Lopez Droguett et al. (2022) [82] developed a full pipeline for crack segmentation on real bridge data. The process started with image acquisition from inspection videos under varied conditions. Then, based on the videos, two datasets were created: a manually labeled set (CRACK_V1) and a larger semi-automatically labeled set (CRACK_V2). The frames were divided into 96×96 patches, with CRACK_V2 additionally pre-processed using gradient filtering, thresholding, and morphology to enhance the visibility of cracks.

Feature extraction relied on deep networks rather than handcrafted features, with architectures such as SegNet and DeepLabV3+ tested, and a lightweight DenseNet-13 introduced to maximize feature reuse. Crack detection was formulated as a semantic segmentation task, with DenseNet-13 employing downsampling and upsampling paths to retain detail while keeping the parameter count low. No complex post-processing was added; instead, class imbalance was addressed using data augmentation and dropout. Results showed DenseNet-13 achieved 94.51% IOU and over 97% accuracy, outperforming much deeper models while using only 350k parameters. With inference times of ~0.001 ms per patch, the approach is suitable for mobile devices and real-time inspection. However, limitations remain, including the field's robustness to shadows and dirt.

6.5.2 Case study ii: Capacity prediction using GA-enhanced SVM

Zhou et al. (2024) [99] developed a GA-SVM framework for bridge safety evaluation that integrates simulated and field data. First, they generated 45 samples from finite element simulations of an RC hollow-slab bridge under different loading and cracking conditions and later validated the model with field measurements from two real bridges. The outputs were transformed into five normalized features: crack damage ratios at three span positions, maximum dynamic deflection, and crack opening ratio, which serve as structural response indices. These features were then used in a support vector regression model with an RBF kernel, where the penalty factor and kernel width were optimized using a genetic algorithm to avoid local minima. To improve interpretability, SHAP analysis quantified the influence of each feature, revealing the crack opening ratio as the most critical factor.

The GA-SVM achieved high prediction accuracy ($R^2 > 0.96$) for both load capacity and serviceability limit states, outperforming ANN baselines, while reducing computational cost from hours of FEM simulation to seconds. However, the method is constrained by the difficulty of accurately measuring crack indices in practice and the small size of the training database.

6.5.3 Case study iii: Portable computer-vision-based load rating

Dong et al. (2020) [111] proposed a portable, vision-based approach for load rating of prestressed concrete bridges. Data were acquired from both controlled load tests (static truck loading) and normal traffic, with video recording of girder deflections. Preprocessing involved camera calibration to convert pixels to real-world units and extraction of image features using SIFT feature detectors and VGG descriptors, followed by feature matching and RANSAC filtering to remove false matches. The displacement of girders was estimated from tracked features, and distribution factors were then computed as the ratio of the deflection of each girder to the sum across all girders. These values were used to calculate load rating factors according to AASHTO specifications.

Analysis showed that the proposed method produced deflections and distribution factors highly consistent with those from strain gauges and calibrated FEM models, but less conservative than AASHTO formulas, which tended to over-allocate load to exterior girders. Results indicated that the proposed approach could improve load rating factors by up to 12% compared with AASHTO estimates, while providing results that are close to those of FEM-based methods. The system achieved this with only three low-cost cameras and without requiring traffic closures. However, performance depends on image quality, environmental factors (such as lighting, occlusion, wind, and vibration), and the presence of sufficient surface texture or artificial markers for feature tracking.

Table 7. Summary of selected case studies applying AI and computer vision for bridge condition assessment and load prediction

Case study	Process	Key results	Efficiency & deployment
i [82]	-Acquisition: The authors captured photographs of concrete bridge surfaces to document crack patterns. -Preprocess: They created pixel-level masks for two newly compiled crack-segmentation datasets and included a public benchmark dataset; pre-processed the data using gradient filtering, thresholding, and morphology to highlight cracks -Model: A lightweight DenseNet-13 segmentation network with one feature extractor and two parallel data paths was trained. -Validate: The model was benchmarked against several state-of-the-art segmentation networks. -Deploy: The architecture was designed to be efficient enough for direct use on mobile devices.	Achieved an IoU of 94.51% for crack segmentation, outperforming baseline algorithms on the same datasets.	Only 350k parameters, making it much smaller than typical segmentation models and feasible for on-device use.
ii [99]	-Acquisition: Static and dynamic simulations of reinforced concrete bridges were run using ANSYS. -Preprocess: Key features such as maximum deflection and crack-opening ratios at mid-, quarter-, and three-quarter spans were extracted. -Model: A support vector machine tuned with a GA was developed -Validate: Performance was compared with an ANN - Interpret: SHAP analysis was applied to identify the most influential features.	GA-SVM outperformed ANN for both load-carrying and serviceability limit predictions; crack-opening ratio was the top predictor.	Small, simulation-based dataset allowed fast training and inference.
iii [111]	-Acquisition: Portable cameras were used to record live traffic on a prestressed concrete highway bridge. -Preprocess: A computer-vision pipeline extracted girder deflections from the video footage. -Model: Calculated live-load distribution factors from the deflection data. Rate: Used these factors to compute bridge load-rating factors. -Validate: Compared results with AASHTO simplified formula, strain-gauge measurements, and a calibrated finite-element model.	Rating factors closely matched those from strain gauges and FEM, and improved by ~12% over the AASHTO simplified formula.	Portable, non-contact setup avoided traffic closures and heavy equipment.

The discussions in Section 6 highlight the transformative potential of AI in concrete bridge health monitoring. AI methods have demonstrated superior accuracy, efficiency, and robustness compared to traditional approaches in detecting concrete damage and predicting structural capacity. However, certain limitations remain as follows:

- SHM of concrete bridge often involves combining datasets from vibration signals, strain measurements, acoustic emissions, and image-based inspections, each with distinct sampling rates, resolutions, and noise characteristics. Integrating these heterogeneous datasets into a coherent diagnostic framework remains a significant technical challenge (see Section 7.1).
- Many AI models are often trained on datasets that suffer from issues such as being small, imbalanced, or containing missing values due to sensor failures or inconsistent monitoring protocols. In SHM, the issues can be amplified by the difficulty of collecting long-term, continuous monitoring data across diverse bridge types, ages, and traffic conditions. The scarcity of representative SHM datasets can limit the statistical robustness of AI models and can introduce bias in predictions (see Section 7.2).
- Deep learning methods have shown strong predictive performance in detecting anomalies or predicting damage progression. However, their black-box nature makes it difficult for engineers to trace how specific sensor readings or structural features drive predictions. This lack of interpretability can undermine trust among SHM practitioners and hinder decision-making in safety-critical infrastructure management (see Section 7.3).
- AI models often generalize poorly when applied to bridge types, structural configurations, or environmental conditions different from those in the training data. The lack of sufficient generalizability is especially problematic in SHM, where models are expected to adapt to diverse environments and evolving deterioration processes (see section 7.4).
- Implementing AI-based SHM in operational practice faces barriers such as the cost and durability of sensor installations, the need for reliable computing infrastructure, and the computational inefficiency of current deep learning models. These constraints can hinder large-scale or real-time deployment (see Section 7.5).

7 Future Directions for AI in Bridge Assessment

To move from experimental studies toward scalable and trustworthy practice, future research is advised to focus on overcoming the five key limitations listed in the end of Section 6. The following subsections outline potential strategies and research opportunities regarding these limitations.

7.1 Data Heterogeneity in Bridge SHM

Heterogeneity in bridge SHM arises in scenarios in which information comes from different sensing modalities or different bridges with varied layouts, materials, and data-collection settings. Several researchers have attempted to alleviate these issues through various approaches. For example, Anterrieu et al. (2019) [101] fused multiple non-destructive testing measurements collected over the same deck, which consisted of ground-penetrating radar, electrical capacitance, half-cell potential, and resistivity. Researchers first aligned all measurements to a common grid. They then assimilated the heterogeneous inputs to predict corrosion rate and its uncertainty, using geostatistical models and a decision-tree baseline for comparison. This approach lets engineers read one coherent corrosion map rather than four separate, hard-to-compare surveys. It handles modality differences by translating every input into a single spatial field with quantified uncertainty.

Ichi and Dorafshan (2022) [112] addressed cross-bridge heterogeneity in UAV infrared thermography for deck delamination. Researchers standardized inputs with semantic segmentation of infra thermography images, so the model pays attention only to the deck region. Then, they optimized detection parameters per bridge cluster, acknowledging that thermal patterns differ across sites. The study reports cluster-specific thresholds and analyses the effect of ambient conditions such as wind. This design groups bridges with similar characteristics together, improving practicality when datasets comprise multiple structures.

Dong et al. (2020) [111] tackled cross-source heterogeneity between vision-based deflection measurements and conventional inspection data. They checked consistency by comparing camera-derived deflections and the resulting load-rating factors against strain-gauge measurements, a calibrated FEM, and AASHTO formulas. Consistency checks align outputs from heterogeneous sources, providing confidence when one source is unavailable. The field study showed similar load-distribution and rating factors across methods.

Table 8 presents the strategies adopted in the studies mentioned above in a structured way, detailing the steps together with the outcomes and limitations, to help readers better understand the way heterogeneous SHM data can be managed in practice.

Table 8. Strategies for handling heterogeneous SHM datasets, organized by data problem, strategy, pipeline step, outcome, and limitations.

Reference	Heterogeneity issue	Strategy used	Step in the pipeline	Outcome
[101]	Mixed NDT modalities over the same deck (different physics/units).	Geostatistical data assimilation (cokriging, Bayesian Sequential Simulation) with a decision-tree baseline.	-Acquire: Ground penetrating radar, electrical capacitance, half-cell potential, and resistivity. -Align: Put all readings on a common grid. -Model: Fit spatial correlation (variograms) and choose assimilation method. -Fuse: Assimilate all NDT to produce a corrosion-rate (Vcor) map with uncertainty; compare against the decision tree.	One coherent corrosion map from heterogeneous NDT with quantified uncertainty, enabling joint use of modalities for diagnosis and planning.
[112]	Cross-bridge heterogeneity (five in-service bridges; varying conditions/layouts).	Standardize inputs + per-group parameter optimization for detection.	-Acquire: UAV infrared thermography over five bridges. -Segment: Semantically segment images so the model focuses on the deck region. -Optimize: Tune thresholds per bridge cluster (2-clustered and 3-clustered groupings) to reflect site differences. -Evaluate: Report accuracy and analyze environmental effects (e.g., wind, defect depth).	Cluster-specific settings that respect differences between bridges; more realistic deployment across sites than one global threshold.
[111]	Cross-source heterogeneity (vision vs. strain gauges vs. FEM/code formulas).	Cross-source consistency checks to align outputs from different modalities.	-Acquire: Portable camera videos under live traffic. -Extract: Use computer vision to compute girder deflections. -Compute: Derive LLDFs and load-rating factors. -Validate: Compare with strain-gauge data, a calibrated FEM, and AASHTO formulas.	Shows agreement between CV-based and traditional sources; provides confidence when one source is missing or impractical (no traffic closure).

7.2 Insufficient and Inconsistent Data

Generative adversarial networks are developed to tackle problems concerning limited labeled and imbalanced data in supervised learning. In addition to its capability of dealing with an insufficient amount of training data, a transfer learning-based approach has been used to reduce the computational effort that results from the huge number of parameters of deep learning models [42].

Moreover, data inconsistency exists between laboratory-based data and field data [113]. The presence of these inconsistencies frequently results in imprecise forecasts of conditions and ineffective design of maintenance strategies. Conventional maintenance management systems inherently lack the essential tools, such as mechanistic solutions and advanced data handling capabilities, required to tackle these inconsistencies adequately. The increasing demand necessitates the development of self-learning systems within maintenance management systems that can effectively convert laboratory-based research findings into practical applications in the field.

Moreover, data inconsistency exists between laboratory-based data and field data [113]. The presence of these inconsistencies frequently results in imprecise forecasts of conditions and ineffective design of maintenance strategies. Conventional maintenance management systems inherently lack the essential tools, such as mechanistic solutions and advanced data handling capabilities, required to tackle these inconsistencies adequately. The increasing demand necessitates the development of self-learning systems within maintenance management systems that can effectively convert laboratory-based research findings into practical applications in the field.

7.3 Interpretability of AI Models

As previously discussed, AI is a powerful tool for predicting, classifying, and extracting features based on a vast amount of data. However, the data cannot explain everything, especially parameters related to complicated structural behavior, which could be dimension-dependent, failure type-dependent, and structural type-dependent. A lack of underlying physics will limit the trustworthiness and accuracy of AI results [95, 96].

Physics and AI hybrid models could be used to solve engineering problems dealing with complex structural behavior and a large amount of data. Several works have applied neural networks and other soft-computing techniques to deal with uncertainty and explain the correlation between model predictions and the physical problem [18]. This helps to improve the interpretability of AI methodologies, particularly neural networks. In this regard, researchers in [17] have recently proposed a new hybrid approach to infuse micromechanics knowledge into the learning process of neural networks. In their study, micromechanical principles were embedded directly into the neural network's loss function, allowing the model to learn from experimental data while simultaneously adhering to physically based constraints. This micromechanics-infused neural network (MINN) framework improved prediction accuracy, interpretability, and robustness compared to purely data-driven models. While this approach was developed in the context of pavement engineering for predicting asphalt mixture stiffness, the same concept can be adapted to concrete bridge assessment by embedding structural mechanics or fracture mechanics models into AI frameworks for predicting sKPIs such as load capacity, crack propagation rates, or shear strength.

Embedding physical constraints may not always be feasible due to data limitations or the absence of well-established analytical models. Therefore, ensuring an in-depth understanding of the decision-making process of purely data-driven models remains crucial, particularly when considering the potential impact on critical infrastructure applications.

Interpretability tools like Local Interpretable Model-agnostic Explanations (LIME) [114] and SHapley Additive exPlanations (SHAP) [115] are important for elucidating the outputs of AI models. These tools evaluate the decision-making process of ML models, enabling engineers to obtain clear insights into the predictive behaviors of the models [116]. Incorporating such interpretability methods into AI frameworks bridges the gap between purely data-driven predictions and the underlying physical principles [117].

7.4 Robustness in Dynamic Environments

A major limitation of AI in bridge assessment is its adaptability to dynamic environments. To overcome this limitation, future research may need to focus on developing adaptive learning frameworks that can accommodate the variability inherent in real-world conditions. AI models must be capable of real-time learning and adaptation to shifting operational and environmental conditions, including climate-induced deterioration, variable loading patterns, and evolving material properties. Moreover, incorporating advanced techniques such as graph signal processing [118] can enhance spatial and temporal data integration, enabling AI systems to operate reliably. Additionally, reinforcement learning can enhance AI robustness by allowing the models to learn optimal decision-making strategies based on continuous feedback. This can improve model resilience in unpredictable scenarios like extreme weather conditions or unforeseen structural damage. Another promising approach is domain adaptation, where AI models trained on data from one environment can be adjusted to new environments with minimal retraining. This is particularly important for bridge monitoring systems deployed in diverse geographical locations. Ensemble learning methods, which combine multiple AI models, can further enhance system resilience by reducing the impact of individual model weaknesses.

Similar challenges in AI robustness have been observed in railway infrastructure, where varying operational loads, environmental conditions, and maintenance practices influence predictive accuracy. Future opportunities are discussed in [119] about AI applications in railway infrastructure that can be leveraged to inform future developments in bridge assessment, fostering cross-domain advancements in AI-driven structural health monitoring.

7.5 Computational Efficiency and Deployment

Computational inefficiencies are a significant limitation of current AI methods, such as resource-intensive models like CNN and semantic segmentation frameworks. Such inefficiencies might not only limit the scalability of AI systems but also hinder their adoption in real-time monitoring scenarios. Addressing this limitation may require research into lightweight AI architectures that reduce computational burdens without compromising accuracy. Techniques such as model pruning [120] and quantization [121] are particularly promising as they streamline the computational demands of deep learning models while preserving their accuracy. Additionally, integrating edge computing [122] solutions for bridge health monitoring systems can decentralize data processing, enabling real-time anomaly detection on-site.

In addition to conventional edge computing, edge-AI [123] integrates trained AI models within on-site processing units. This integration minimizes data transmission needs by performing inference at the source of the data, thereby reducing decision-making latency. For latency-aware SHM applications, such as vibration-based anomaly detection or real-time tracking of crack propagation, edge-AI can provide near-instant alerting without relying on ubiquitous high-bandwidth connectivity.

8 Conclusion

Assessment of existing concrete bridges is crucial for maintaining the safety and functionality of aging infrastructure. Traditional methods for bridge assessment rely on manual inspections and physics-based models, which are often time-consuming and costly. Given the increasing availability of monitoring data and advancements in AI, this review was conducted to explore the role of AI in bridge assessment, specifically for real concrete bridges, identifying both its benefits and challenges.

A literature review was conducted using the latest 224 papers from Scopus, selected based on relevance to AI applications in real-world concrete bridge monitoring. The papers were categorized according to the data type processed, image, signal, or point data, and the sKPIs they addressed. AI models were evaluated in terms of accuracy, efficiency, interpretability, and robustness in real-world conditions. The review also examined the limitations of existing AI models and explored potential enhancements, including physics-informed AI approaches and uncertainty quantification.
The key findings from our review are listed below.

- Choice of AI techniques is strongly aligned with specific sKPIs. Image-based methods (CNNs, U-Nets, YOLO) are most effective for crack detection and segmentation, achieving high accuracy but requiring large, annotated datasets. Signal-based methods (such as SVM) are suited for vibration and time-series analysis. Ensemble learning and regression models are effective for capacity and deflection estimation, especially when optimized with meta-heuristics (e.g., GA-SVM).
- Robustness in real-world environments remains a challenge. Models trained only on clean, lab-generated data lose accuracy under field conditions with shadows, stains, or noise. Data augmentation and hybrid pipelines (e.g., FCN + naive Bayes, shadow removal filters) significantly improve robustness for practical deployment.
- Embedding physical constraints (e.g., structural mechanics, micromechanics) into learning models improves reliability, interpretability, and generalization compared to purely data-driven methods. And hybrid updating with FEM and Bayesian methods helps align predictions with actual structural behaviour.
- Effective fusion of multi-modal data (e.g., GPR, infrared thermography, strain sensors) enables more comprehensive and reliable condition assessment. And feature-level fusion (e.g., HOG + LBP descriptors for cracks) has proven superior to single-modality features, improving classification accuracy.
- Computationally heavy models (e.g., deep CNNs, segmentation networks) pose challenges for real-time SHM. While lightweight AI and edge-AI frameworks reduce latency, enabling real-time monitoring.

In general, this review highlights that while AI has demonstrated significant potential in enhancing the accuracy and efficiency of bridge assessment, its widespread adoption in practice still faces barriers related to robustness, interpretability, and data scarcity. By integrating physics-informed approaches, advancing uncertainty quantification, and developing scalable frameworks for multi-modal data fusion, AI can evolve from promising experimental applications to reliable tools for large-scale, real-world deployment. The findings of this study, therefore, are expected to provide a foundation for guiding both future research and practical implementation of AI-driven strategies in the structural assessment and long-term management of concrete bridges.

9 CRediT authorship contribution statement

Mohammadjavad Berangi: Methodology, Investigation, Writing (Review and Editing), Visualization. **Fengqiao Zhang:** Conceptualization, Methodology, Investigation, Writing (Original Draft), Visualization. **Wassamon Phusakulkajorn:** Methodology, Investigation, Writing (Review and Editing). **Alfredo Núñez:** Funding, Conceptualization, Supervision, Writing (Review and Editing). **Kumar Anupam:** Funding, Conceptualization, Supervision, Writing (Review and Editing).

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of the manuscript, the authors used Grammarly and Quillbot to improve the language and readability of the paper. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to express their sincere gratitude to colleagues at the Department of Engineering Structures at Delft University of Technology for their insightful discussions and technical support throughout the development of this work. Special thanks are extended to those who assisted with manuscript proofreading and provided valuable feedback on the clarity and organization of the content.

Data availability statements

The authors declare that the data supporting the findings of this study are available within the paper.

Reference

1. Bleijenberg, A.N., *Renewal of civil infrastructure: Dutch national forecast for replacement and renovation*. 2021, TNO: Delft, the Netherlands. p. 39.
2. Srikanth, I. and M. Arockiasamy, *Remaining Service Life Prediction of Aging Concrete Bridges Based on Multiple Relevant Explanatory Variables.* Practice Periodical on Structural Design and Construction, 2021. **26**(4): p. 04021036.
3. Rijkswaterstaat, *Staat van de Infra*. 2021.
4. Rijkswaterstaat, *Vervanging en Renovatie: Prognose voor de periode 2023 tot en met 2050*, in *forecast report*. 2022: the Netherlands.
5. ASCE, *A Comprehensive Assessment of America's Infrastructure*. 2021.
6. Rijkswaterstaat, *Inventarisatie kunstwerken (bruggen, tunnels en viaducten)*. 2007. p. 47.
7. *fib, Monitoring and safety evalutation of eixisting cocrete structures*. *fib* bulletin. Vol. 22. 2003: Elsener and Böhni. 297.
8. Belletti, B., et al., *Analytical and numerical evaluation of the design shear resistance of reinforced concrete slabs.* Structural Concrete, 2014. **15**(3): p. 317-330.
9. Strauss, A., et al., *Performance indicators for road bridges - overview of findings and future progress*. 2017.
10. Hassani, S. and U. Dackermann, *A Systematic Review of Advanced Sensor Technologies for Non-Destructive Testing and Structural Health Monitoring.* Sensors, 2023. **23**(4): p. 2204.
11. Sun, L., et al., *Review of Bridge Structural Health Monitoring Aided by Big Data and Artificial Intelligence: From Condition Assessment to Damage Detection.* Journal of Structural Engineering, 2020. **146**(5): p. 04020073.
12. Goodfellow, I., Y. Bengio, and A. Courville, *Deep Learning*. 2016: MIT Press.
13. Di Mucci, V.M., et al., *Artificial intelligence in structural health management of existing bridges.* Automation in Construction, 2024. **167**: p. 105719.
14. Miao, Z., et al., *Pixel-level multicategory detection of visible seismic damage of reinforced concrete components.* Computer-Aided Civil and Infrastructure Engineering, 2021. **36**(5): p. 620-637.
15. Hou, R. and Q. Hou, *Prediction of the shear capacity of ultrahigh-performance concrete beams using neural network and genetic algorithm.* Scientific Reports, 2023. **13**(1): p. 2145.
16. Rudin, C., *Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead.* Nature Machine Intelligence, 2019. **1**(5): p. 206-215.
17. Anupam, K., et al., *A step toward a micromechanics-informed neural network for predicting asphalt mixture stiffness.* Computer-Aided Civil and Infrastructure Engineering, 2025. **40**(23): p. 3624-3651.
18. Phusakulkajorn, W., et al., *Spiking neural network with time-varying weights for rail squat detection.* Applied Soft Computing, 2025. **184**: p. 113689.
19. Xu, D., et al., *Concrete and steel bridge Structural Health Monitoring—Insight into choices for machine learning applications.* Construction and Building Materials, 2023. **402**: p. 132596.
20. Prakash, V., et al., *Structural Health Monitoring of Concrete Bridges Through Artificial Intelligence: A Narrative Review.* Applied Sciences, 2025. **15**(9): p. 4855.
21. Moher, D., et al., *Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement.* International Journal of Surgery, 2010. **8**(5): p. 336-341.
22. Adeli, H., *Neural Networks in Civil Engineering: 1989–2000.* Computer-Aided Civil and Infrastructure Engineering, 2001. **16**(2): p. 126-142.
23. Ko, J.M. and Y.Q. Ni, *Technology developments in structural health monitoring of large-scale bridges.* Engineering Structures, 2005. **27**(12): p. 1715-1725.
24. Lin, W. and T. Yoda, *Chapter Six - Reinforced and Prestressed Concrete Bridges*, in *Bridge Engineering*, W. Lin and T. Yoda, Editors. 2017, Butterworth-Heinemann. p. 97-110.
25. Lin, W. and T. Yoda, *Reinforced and Prestressed Concrete Bridges*. 2017. p. 97-110.
26. Park, R. and W.L. Gamble, *Reinforced Concrete Slabs*. 1999: Wiley.
27. Deutscher Ausschuss für Stahlbeton, *DAtStb-Guideline: Load Tests on Concrete Structures (in German) DAfStb-Richtlinie : Belastungsversuche an Betonbauwerken*, D.A.f. Stahlbeton, Editor. 2000: Berlin.
28. ACI Committee 437, *Code requirenments for Load Testing of Existing Concrete Structures (ACI 437.2M-13) and Commentary*. 2013, American Concrete Institute: Farmington Hills, MI.

29. Dieteren, G., et al., *Development of a Smart Key Performance Indicator for In-Situ Load Tests*, in *fib symposium*. 2017: Maastricht.
30. Benitez, K., E.O.L. Lantsoght, and Y. Yang, *Development of a Stop Criterion for Load Tests based on the Critical Shear Displacement Theory*, in *IALCCE*. 2018: Gent, Belgium.
31. Zhang, F., *Acoustic emission-based indicators of shear failure of reinforced concrete structures without shear reinforcement*. 2022, Delft University of Technology: the Netherlands.
32. Collins, M.P. and D. Mitchell, *Prestressed Concrete Structures*. 1991: Prentice Hall.
33. Hewson, N.R., *Prestressed concrete bridges: design and construction*. 2003: Thomas Telford.
34. Roosen, M., et al., *Resistance to diagonal tension cracking of single span prestressed girders*, in *SEMC 2019*. 2019: Cape Town.
35. Zarate Garnica, G., et al., *Monitoring structural responses during proof load testing of reinforced concrete bridges: A review*, in *Tenth International Conference on Bridge Maintenance, Safety and Management (IABMAS 2020)*. 2021: online.
36. Yang, Y., J. den Uijl, and J. Walraven, *Critical shear displacement theory: on the way to extending the scope of shear design and assessment for members without shear reinforcement.* Structural Concrete, 2016. **17**(5): p. 790-798.
37. Tung, N.D. and N.V. Tue, *Effect of support condition and load arrangement on the shear response of reinforced concrete beams without transverse reinforcement.* Engineering Structures, 2016. **111**: p. 370-382.
38. Crisci, G., et al., *Performance Of Existing Reinforced Concrete Arch Bridges Under Current non Seismic Loads*. 2021.
39. Zhang, Z., *Missing data imputation: focusing on single imputation.* Annals of Translational Medicine, 2016. **4**(1): p. 9.
40. Shorten, C. and T. Khoshgoftaar, *A survey on Image Data Augmentation for Deep Learning.* Journal of Big Data, 2019. **6**.
41. Rahm, E. and H.H. Do, *Data cleaning: Problems and current approaches.* IEEE Data Eng. Bull., 2000. **23**(4): p. 3-13.
42. Shorten, C. and T.M. Khoshgoftaar, *A survey on Image Data Augmentation for Deep Learning.* Journal of Big Data, 2019. **6**(1): p. 60.
43. Buda, M., A. Maki, and M.A. Mazurowski, *A systematic study of the class imbalance problem in convolutional neural networks.* Neural Networks, 2018. **106**: p. 249-259.
44. Bhaya, W., *Review of Data Preprocessing Techniques in Data Mining.* Journal of Engineering and Applied Sciences, 2017. **12**: p. 4102-4107.
45. Zheng, A. and A. Casari, *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*. 2018: O'Reilly.
46. Zeng, Y., et al., *Spectrum Analysis and Convolutional Neural Network for Automatic Modulation Recognition.* IEEE Wireless Communications Letters, 2019. **8**: p. 929-932.
47. Zoubir, H., et al., *Concrete Bridge Crack Image Classification Using Histograms of Oriented Gradients, Uniform Local Binary Patterns, and Kernel Principal Component Analysis.* Electronics (Switzerland), 2022. **11**(20).
48. Noori Hoshyar, A., et al., *Algorithm Development for the Non-Destructive Testing of Structural Damage.* Applied Sciences, 2019. **9**(14): p. 2810.
49. Ye, X., et al., *An Integrated Machine Learning Algorithm for Separating the Long-Term Deflection Data of Prestressed Concrete Bridges.* Sensors, 2018. **18**(11): p. 4070.
50. Akintunde, E., et al., *Unsupervised Machine Learning for Robust Bridge Damage Detection: Full-Scale Experimental Validation.* Engineering Structures, 2021. **249**: p. 113250.
51. Kim, B. and S. Cho, *Image-based concrete crack assessment using mask and region-based convolutional neural network.* Structural Control and Health Monitoring, 2019. **26**(8): p. e2381.
52. Islam, M.M., et al., *CNN Based on Transfer Learning Models Using Data Augmentation and Transformation for Detection of Concrete Crack.* Algorithms, 2022. **15**(8): p. 287.
53. Jolliffe, I.T. and J. Cadima, *Principal component analysis: a review and recent developments.* Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2016. **374**(2065): p. 20150202.
54. Schölkopf, B., A. Smola, and K.-R. Müller. *Kernel principal component analysis*. in *International conference on artificial neural networks*. 1997. Springer.
55. Hinton, G.E. and R.R. Salakhutdinov, *Reducing the Dimensionality of Data with Neural Networks.* Science, 2006. **313**(5786): p. 504-507.
56. Islam, M.M.M. and J.-M. Kim, *Vision-Based Autonomous Crack Detection of Concrete Structures Using a Fully Convolutional Encoder–Decoder Network.* Sensors, 2019. **19**(19): p. 4251.
57. Jain, A.K., *Data clustering: 50 years beyond K-means.* Pattern Recognition Letters, 2010. **31**(8): p. 651-666.
58. MacQueen, J. *Some methods for classification and analysis of multivariate observations*. in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. 1967. Oakland, CA, USA.
59. Johnson, S.C., *Hierarchical clustering schemes.* Psychometrika, 1967. **32**(3): p. 241-254.
60. Murtagh, F. and P. Legendre, *Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?* Journal of Classification, 2014. **31**(3): p. 274-295.
61. Bishop, C.M., *Pattern recognition and machine learning.* Springer google schola, 2006. **2**: p. 5-43.
62. Cortes, C. and V. Vapnik, *Support-vector networks.* Machine Learning, 1995. **20**(3): p. 273-297.

63. LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning.* Nature, 2015. **521**(7553): p. 436-444.
64. Quinlan, J.R., *Induction of decision trees.* Machine Learning, 1986. **1**(1): p. 81-106.
65. Breiman, L., *Random Forests.* Machine Learning, 2001. **45**(1): p. 5-32.
66. Tan, C., et al. *A survey on deep transfer learning*. in *Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III 27*. 2018. Springer.
67. Guzmán-Torres, J.A., M.Z. Naser, and F.J. Domínguez-Mota, *Effective medium crack classification on laboratory concrete specimens via competitive machine learning.* Structures, 2022. **37**: p. 858-870.
68. Garcia-Garcia, A., et al., *A review on deep learning techniques applied to semantic segmentation.* arXiv preprint arXiv:1704.06857, 2017.
69. James, G., et al., *An Introduction to Statistical Learning: with Applications in R*. 2013: Springer New York.
70. Kochenderfer, M.J. and T.A. Wheeler, *Algorithms for optimization*. 2019: Mit Press.
71. Goldberg, D.E., *Genetic Algorithms in Search, Optimization and Machine Learning*. 1989: Addison-Wesley Longman Publishing Co., Inc.
72. Otsu, N., *A Threshold Selection Method from Gray-Level Histograms.* IEEE Transactions on Systems, Man, and Cybernetics, 1979. **9**(1): p. 62-66.
73. Li, Y., H. Li, and H. Wang, *Pixel-wise crack detection using deep local pattern predictor for robot application.* Sensors (Switzerland), 2018. **18**(9).
74. Nomura, Y., M. Inoue, and H. Furuta, *Evaluation of crack propagation in concrete bridges from vehicle-mounted camera images using deep learning and image processing.* Frontiers in Built Environment, 2022. **8**.
75. Kumar, S.S., et al., *Automated defect classification in sewer closed circuit television inspections using deep convolutional neural networks.* Automation in Construction, 2018. **91**: p. 273-283.
76. Zoubir, H., et al., *Concrete Bridge Crack Image Classification Using Histograms of Oriented Gradients, Uniform Local Binary Patterns, and Kernel Principal Component Analysis.* Electronics, 2022. **11**(20): p. 3357.
77. Modarres, C., et al., *Convolutional neural networks for automated damage recognition and damage type identification.* Structural Control and Health Monitoring, 2018. **25**(10): p. e2230.
78. Pozzer, S., et al., *Semantic Segmentation of Defects in Infrared Thermographic Images of Highly Damaged Concrete Structures.* Journal of Performance of Constructed Facilities, 2021. **35**(1): p. 04020131.
79. Liu, Z., et al., *Computer vision-based concrete crack detection using U-net fully convolutional networks.* Automation in Construction, 2019. **104**: p. 129-139.
80. Adel, M., et al., *Early damage detection of fatigue failure for RC deck slabs under wheel load moving test using image analysis with artificial intelligence.* Engineering Structures, 2021. **246**: p. 113050.
81. Venkataramani, S., et al., *Efficient AI System Design With Cross-Layer Approximate Computing.* Proceedings of the IEEE, 2020. **108**(12): p. 2232-2250.
82. Lopez Droguett, E., et al., *Semantic segmentation model for crack images from concrete bridges for mobile devices.* Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 2022. **236**(4): p. 570-583.
83. Ren, S., et al., *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017. **39**(6): p. 1137-1149.
84. Zhang, J., S. Qian, and C. Tan, *Automated bridge surface crack detection and segmentation using computer vision-based deep learning model*. Engineering Applications of Artificial Intelligence, 2022. **115**: p. 105225.
85. Lee, S., et al., *Deep Learning-Based PC Member Crack Detection and Quality Inspection Support Technology for the Precise Construction of OSC Projects.* Applied Sciences, 2022. **12**(19): p. 9810.
86. Matsushita, Y., et al., *Illumination normalization with time-dependent intrinsic images for video surveillance.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004. **26**(10): p. 1336-1347.
87. Pal, M., et al., *An Overview of Challenges Associated with Automatic Detection of Concrete Cracks in the Presence of Shadows.* Applied Sciences, 2021. **11**(23): p. 11396.
88. Li, G., et al., *Recognition and evaluation of bridge cracks with modified active contour model and greedy search-based support vector machine.* Automation in Construction, 2017. **78**: p. 51-61.
89. Li, G., et al., *Automatic crack recognition for concrete bridges by fully convolutional neural network and Naive Bayes data fusion based on visual detection system.* Measurement Science and Technology, 2020. **31**.
90. Srikanth, I. and M. Arockiasamy, *Deterioration models for prediction of remaining useful life of timber and concrete bridges: A review.* Journal of Traffic and Transportation Engineering (English Edition), 2020. **7**(2): p. 152-173.
91. Chan, T.H.T., J. Li, and C. Caprani, *Structural identification and evaluation for SHM applications.* Journal of Civil Structural Health Monitoring, 2018. **8**(5): p. 719-720.
92. Azizzadenesheli, K., et al., *Neural operators for accelerating scientific simulations and design.* Nature Reviews Physics, 2024: p. 1-9.
93. Stoffel, M., F. Bamer, and B. Markert, *Artificial neural networks in structural dynamics.* PAMM, 2019. **19**(1): p. e201900128.
94. Kumar, A., et al., *Machine learning intelligence to assess the shear capacity of corroded reinforced concrete beams.* Scientific Reports, 2023. **13**(1): p. 2857.
95. Fathalla, E., Y. Tanaka, and K. Maekawa, *Remaining fatigue life assessment of in-service road bridge decks based upon artificial neural networks.* Engineering Structures, 2018. **171**: p. 602-616.

96. Lattanzi, D., et al., *Bridge Column Maximum Drift Estimation via Computer Vision.* Journal of Computing in Civil Engineering, 2016. **30**(4): p. 04015051.
97. Davoudi, R., G.R. Miller, and J.N. Kutz, *Structural Load Estimation Using Machine Vision and Surface Crack Patterns for Shear-Critical RC Beams and Slabs.* Journal of Computing in Civil Engineering, 2018. **32**(4): p. 04018024.
98. Jiang, C.-S., et al., *Hybrid genetic algorithm and support vector regression for predicting the shear capacity of recycled aggregate concrete beam.* Soft Computing, 2024. **28**(2): p. 1023-1039.
99. Zhou, S., D. Yan, and Y. He, *Interpretable Capacity Prediction of RC Bridges Based on Genetic Algorithm-enhanced Support Vector Machine Learning.* KSCE Journal of Civil Engineering, 2024.
100. Locke, W., et al., *An Intelligently Designed AI for Structural Health Monitoring of a Reinforced Concrete Bridge*. 2021.
101. Anterrieu, O., et al., *Non-destructive data assimilation as a tool to diagnose corrosion rate in reinforced concrete structures.* Journal of Building Engineering, 2019. **23**: p. 193-206.
102. Stein, A. and L. Corsten, *Universal kriging and cokriging as a regression procedure.* Biometrics, 1991: p. 575-587.
103. Prasanna, P., et al., *Automated Crack Detection on Concrete Bridges.* IEEE Transactions on Automation Science and Engineering, 2016. **13**(2): p. 591-599.
104. Das, A.K., D. Suthar, and C.K.Y. Leung, *Machine learning based crack mode classification from unlabeled acoustic emission waveform features.* Cement and Concrete Research, 2019. **121**: p. 42-57.
105. Mandal, D.D., et al., *Acoustic Emission Monitoring of Progressive Damage of Reinforced Concrete T-Beams under Four-Point Bending.* Materials, 2022. **15**(10): p. 3486.
106. Kang, D. and Y.-J. Cha, *Autonomous UAVs for Structural Health Monitoring Using Deep Learning and an Ultrasonic Beacon System with Geo-Tagging.* Computer-Aided Civil and Infrastructure Engineering, 2018. **33**(10): p. 885-902.
107. Ni, F., J. Zhang, and Z. Chen, *Zernike-moment measurement of thin-crack width in images enabled by dual-scale deep learning.* Computer-Aided Civil and Infrastructure Engineering, 2019. **34**(5): p. 367-384.
108. Bae, H., K. Jang, and Y.-K. An, *Deep super resolution crack network (SrcNet) for improving computer vision–based automated crack detectability in in situ bridges.* Structural Health Monitoring, 2021. **20**(4): p. 1428-1442.
109. Ma, H., et al., *Parallel Systems for the Bridge Inspection.* IEEE Journal of Radio Frequency Identification, 2022. **6**: p. 783-786.
110. Xu, J.-G., et al., *Data-driven rapid damage evaluation for life-cycle seismic assessment of regional reinforced concrete bridges.* Earthquake Engineering & Structural Dynamics, 2022. **51**(11): p. 2730-2751.
111. Dong, C.-Z., S. Bas, and F.N. Catbas, *A portable monitoring approach using cameras and computer vision for bridge load rating in smart cities.* Journal of Civil Structural Health Monitoring, 2020. **10**(5): p. 1001-1021.
112. Ichi, E. and S. Dorafshan, *Effectiveness of infrared thermography for delamination detection in reinforced concrete bridge decks.* Automation in Construction, 2022. **142**: p. 104523.
113. Berangi, M., et al., *Gradient boosting decision trees to study laboratory and field performance in pavement management.* Computer-Aided Civil and Infrastructure Engineering. **n/a**(n/a).
114. Ribeiro, M.T., S. Singh, and C. Guestrin, *"Why Should I Trust You?": Explaining the Predictions of Any Classifier*, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016, Association for Computing Machinery: San Francisco, California, USA. p. 1135–1144.
115. Lundberg, S.M. and S.-I. Lee, *A unified approach to interpreting model predictions*, in *Proceedings of the 31st International Conference on Neural Information Processing Systems*. 2017, Curran Associates Inc.: Long Beach, California, USA. p. 4768–4777.
116. Molnar, C., *Interpretable Machine Learning, 2nd Edition: A Guide for Making Black Box Models Explainable*. 2020: Independently published (2022); eBook (2nd Edition, 2024-05-26; Creative Commons Licensed).
117. Raissi, M., P. Perdikaris, and G.E. Karniadakis, *Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations.* Journal of Computational Physics, 2019. **378**: p. 686-707.
118. Cheema, M.A., et al., *Computationally Efficient Structural Health Monitoring Using Graph Signal Processing.* IEEE Sensors Journal, 2024. **24**(7): p. 11895-11905.
119. Phusakulkajorn, W., et al., *Artificial intelligence in railway infrastructure: current research, challenges, and future opportunities.* Intelligent Transportation Infrastructure, 2023. **2**.
120. Dai, S., et al., *Energy-Efficient Inference With Software–Hardware Co-Design for Sustainable Artificial Intelligence of Things.* IEEE Internet of Things Journal, 2024. **11**(24): p. 39170-39182.
121. Singh, S., K. Sharma, and B. Karna, *Pruning and Quantization for Deeper Artificial Intelligence (AI) Model Optimization*. 2023. p. 933-945.
122. Gutiérrez-Zaballa, J., K. Basterretxea, and J. Echanobe, *Evaluating single event upsets in deep neural networks for semantic segmentation: An embedded system perspective.* Journal of Systems Architecture, 2024. **154**: p. 103242.
123. Singh, R. and S.S. Gill, *Edge AI: A survey.* Internet of Things and Cyber-Physical Systems, 2023. **3**: p. 71-92.