Progress in Engineering Science 3 (2026) 100308

Contents lists available at ScienceDirect

Progress in Engineering Science

journal homepage: www.sciencedirect.com/journal/progress-in-engineering-science

AI contribution to the monitoring and safety assessment of dams: Review and perspectives[☆]

Caterina Nogara [*], Gabriella Bolzon

Department of Civil and Environmental Engineering, Politecnico di Milano, Piazza Leonardo da Vinci 32, Milan, 20133, Italy

ARTICLE INFO

Keywords:
Dam surveillance
Structural health monitoring
Deep learning
Computer vision
Anomaly detection

ABSTRACT

Structural health monitoring systems are essential to support dam safety assessment, especially as ageing processes and changes in environmental conditions increase uncertainty in the long-term performance of these strategic infrastructures. In recent years, Artificial Intelligence (AI), particularly machine learning, has been increasingly applied to support the interpretation of monitoring data, driven by advances in algorithms as well as improvements in data acquisition and management.

Most existing reviews on AI applications in dam engineering either adopt broad perspectives covering multiple disciplines or focus on specific techniques. In contrast, this article provides a surveillance-oriented narrative review, organizing the literature according to the main components of dam monitoring and the types of data involved, rather than exclusively by algorithm families.

Three complementary areas are considered: continuous monitoring based on point-wise time series, vision-based inspection using RGB images, and full-field non-destructive testing based on optical techniques, with particular attention to infrared thermography and digital image correlation and the role of their outputs in supporting numerical model updating. For each area, the review discusses the surveillance tasks addressed, the types of methods currently adopted, and their main advances and limitations. Finally, open challenges and future research directions are outlined, with a focus on data integration, physical consistency, and interpretability.

1. Introduction

Most concrete dams worldwide are approaching or have exceeded their original design life, while continuing to play a critical role in flood protection, water resources management, and energy production. Over time, these structures are affected by progressive material deterioration processes, such as cracking, expansion phenomena, and surface degradation, which gradually reduce mechanical performance and safety margins [1,2]. At the same time, climate change is modifying environmental actions and operating conditions by altering hydrological loads and thermal regimes, thereby increasing the uncertainty associated with dam response [3]. The combined effects of structural ageing and climate change have been shown to significantly influence the risk of dam failure [4,5].

These challenges reinforce the importance of surveillance in dam safety practice, which combines visual inspections with instrumental measurements and, in some cases, non-destructive and destructive testing, as outlined in international guidelines [6,7]. Continuous instrumental acquisition of environmental actions and structural responses, together with their systematic interpretation, forms the core of Structural Health Monitoring (SHM), a concept widely adopted across different types of civil infrastructure. SHM supports the assessment of structural condition through the analysis and recognition of patterns in measured data, following a hierarchical progression from damage detection to damage identification [8,9].

Within SHM, data-driven pattern-recognition approaches have progressively complemented traditional interpretation methods based on numerical models and classical statistical analyses, with Machine Learning (ML) playing an increasingly central role. Early applications of ML in dam engineering primarily focused on modelling normal structural behaviour from long-term monitoring time series and on supporting anomaly detection through the identification of deviations from expected responses under given environmental and operational conditions [10].

[☆] This article is part of a Special issue entitled: 'PES_IGF28-MedFract3' published in Progress in Engineering Science.

[*] Corresponding author.

E-mail addresses: caterina.nogara@polimi.it (C. Nogara), gabriella.bolzon@polimi.it (G. Bolzon).

https://doi.org/10.1016/j.pes.2026.100308
Received 24 February 2026; Received in revised form 27 April 2026; Accepted 21 May 2026

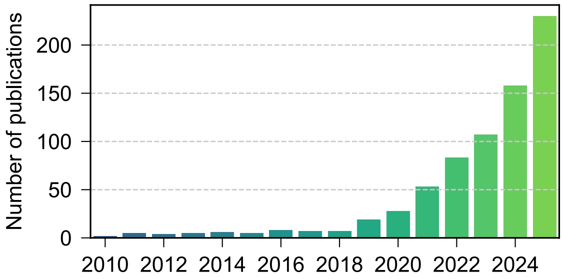

Fig. 1. Number of publications per year retrieved from a Scopus-based search on artificial intelligence applications in dam monitoring.

Recent developments in sensing technologies and data science have driven a rapid increase in research on AI applications in dam monitoring. A Scopus keyword-based search using terms related to dams, monitoring, and artificial intelligence shows a clear growth in contributions, with a marked acceleration after 2020, as illustrated in Fig. 1. Consistently, a keyword co-occurrence analysis of the same dataset highlights the evolving structure of the field, as shown in Fig. 2: machine learning remains central, while Deep Learning (DL) has gained prominence in recent studies, with strong connections to emerging topics such as computer vision, convolutional neural networks, and remote sensing. Both ML and DL are applied to core SHM tasks, including deformation prediction and damage detection. In addition, DL further supports the analysis of more complex data sources, such as UAV-based image data and full-field measurements, complementing traditional monitoring records [11].

Several review papers have examined AI in dam engineering from broad, system-level perspectives. These works typically organize AI applications according to engineering domains and life-cycle stages, covering areas such as hydrological forecasting, energy optimization, design, operation, and risk assessment, often in combination with emerging digital technologies [12,13]. In this context, monitoring and surveillance are treated as one component within a wider technological framework, rather than as a dedicated focus.

Within dam monitoring, earlier review studies are rooted in the traditional distinction between static and dynamic monitoring [14]. However, unlike bridge engineering, where dynamic monitoring is often more informative, static approaches are more frequently adopted in the context of dams, whose structural response is predominantly governed by slowly varying seasonal actions. Building on this foundation, a substantial body of review literature focuses on the modelling of static dam behaviour, where statistical and ML approaches are primarily examined as predictive tools, with emphasis on methodological aspects such as validation and accuracy [15,16]. Other reviews adopt a broader, model-oriented perspective, comparing numerical, statistical, ML, and hybrid approaches within SHM decision-making contexts, and including considerations on non-destructive testing techniques [17,18]. Additional contributions adopt a method-centric organization, analysing ML and DL techniques primarily by algorithm families, with limited distinction among monitoring applications and data types [19].

Complementary to point-wise monitoring, the use of vision-based information has been addressed through a separate body of literature. Reviews explicitly focused on vision-based inspection of dams remain limited, with only a few contributions dedicated to aspects such as image enhancement and concrete defect detection [20]. In contrast, numerous studies addresses visual problems, mainly related to surface damage and crack detection, from a purely algorithmic perspective on generic concrete structures [21,22].

Overall, existing surveys tend to adopt either very broad perspectives, covering AI applications across multiple domains, or highly specific viewpoints focused on individual data sources or methodological aspects. Instead, this paper aims to provide a surveillance-oriented synthesis, focusing on how AI supports key surveillance tasks such as anomaly detection, damage detection and identification, and the interpretation of evolving structural behaviour. The literature is organized according to the three main components of dam surveillance, each associated with a specific type of data: point-wise monitoring time series, visual and imaging data, and full-field measurements providing spatially distributed physical information, such as displacement or temperature fields. These full-field measurements are typically obtained through advanced non-destructive measurement techniques (e.g. digital image correlation and infrared thermography) and complement point-wise monitoring and visual inspection.

After introducing the background on dam monitoring practices and the adopted review approach in Section 2, Section 3 examines ML and DL methods for time-series modelling and anomaly detection. Section 4 focuses on vision-based approaches for surface damage identification, while Section 5 addresses the integration of full-field measurements with AI for damage assessment and model interpretation. Each section concludes with a discussion of current limitations, practical deployment considerations, and future research directions. Finally, Section 6 summarizes the main findings and outlines perspectives for AI-based dam SHM.

2. Background and review methodology

2.1. *Structural health monitoring of dams*

Dam performance is defined as the structural and hydraulic response to operational and environmental loads, including reservoir level variations, thermal actions, and, in some cases, seismic excitation. Surveillance primarily refers to visual inspections aimed at identifying surface anomalies and unexpected changes, whereas monitoring involves the acquisition of quantitative data to assess the response of the structure. Together, they provide complementary information to compare observed behaviour with that expected from design assumptions or predictive models.

Monitoring systems link measured variables to physical processes and, ultimately, to potential failure mechanisms [23]. The variables used to characterize the mechanical response of dams include: (i) displacements, measured through geodetic surveys, pendulums, or inclinometers; (ii) temperature at discrete points, obtained from embedded thermometers or thermistors, as spatial temperature gradients induce deformations and stress redistribution, particularly in arch dams; (iii) stress–strain conditions, measured through strain gauges, describing the internal state of the material; and (iv) joint or crack opening, measured using extensometers or joint metres.

Similarly, hydraulic variables are measured to characterize internal flow processes and their impact on stability, which are particularly important for certain types of dams, such as embankment dams. These include: (i) pore and uplift pressures, measured using piezometers, influencing effective stresses within the dam and its foundation; and (ii) seepage quantities, obtained from flow measurements, providing insight into flow paths and their evolution, potentially indicating internal erosion or leakage.

Traditional dam surveillance systems rely primarily on displacement measurements collected at local instrumented points. Consequently, much of the reviewed literature focuses on predictive modelling and anomaly detection based on these point-wise displacement time series, while hydraulic-related applications are covered less extensively and are often more case-specific.

Complementary investigations include destructive tests and non-destructive techniques, such as ultrasonic testing and infrared thermography, which provide additional insight into both mechanical response and hydraulic conditions.

The effectiveness of surveillance and monitoring is governed by how data are acquired, processed, and interpreted. Visual inspections

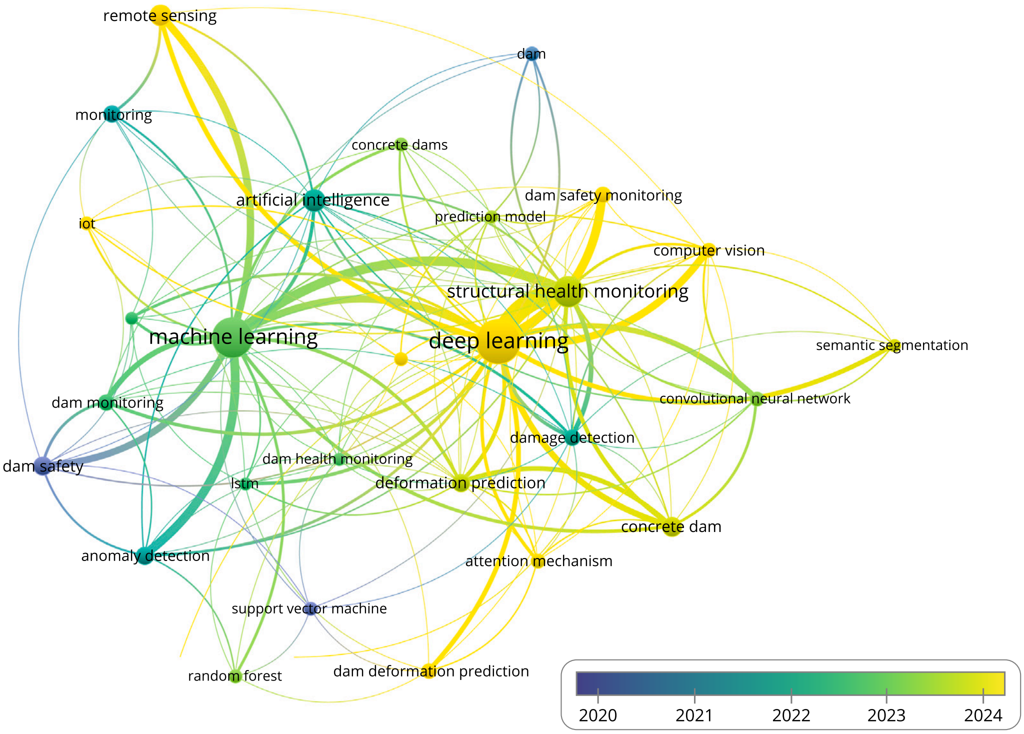

Fig. 2. Keyword co-occurrence network obtained from the same Scopus dataset using VOSviewer. Node size is proportional to keyword frequency, while links indicate co-occurrence relationships.

are conducted periodically and after significant events, while instrumented data are collected manually or automatically at frequencies adapted to each parameter [24]. Measurements are evaluated against expected behaviour and predefined thresholds. Exceedances trigger further investigation, intensified monitoring, or immediate actions. This evaluation–decision–action process is central to dam safety management and is increasingly supported by data-driven approaches to handle complex and heterogeneous datasets.

2.2. Review approach

This review adopts a narrative approach, where literature selection was guided by thematic relevance rather than a fully systematic protocol. The Scopus database was used as the primary tool to support the search process. In this context, the TITLE-ABS-KEY field was queried to retrieve expressions from titles, abstracts, and author keywords.

The search strategy was defined iteratively. An initial query based on general expressions related to dams, monitoring, and artificial intelligence was used as a preliminary screening of the literature. Subsequently, more targeted searches were developed separately for each of the selected dam monitoring domains.

For each domain, the query structure was based on the combination of elements describing the structural context, the monitoring or measurement aspects, data-driven methods, and task-specific objectives. Keywords within each group were connected using OR operators, while different groups were combined using AND operators.

As an example, a representative query for point-wise monitoring combined the following keywords in the TITLE-ABS-KEY field: (dam OR “concrete dam*” OR “arch dam*”) AND (monitoring OR “structural health monitoring” OR SHM OR “dam monitoring” OR “dam safety”) AND (“machine learning” OR “deep learning” OR “data-driven” OR “support vector machine*” OR “random forest” OR “boosted regression tree*”) AND (“anomaly detection” OR prediction OR “displacement prediction” OR “behaviour prediction”).

For vision-based inspection, the search focused on concepts related to image acquisition, such as visual inspection, image processing, remote sensing, and UAVs, and linked them to computer vision techniques including segmentation, object detection, and convolutional neural networks, as well as to application-related keywords such as crack detection and damage identification. The same query structure was applied to full-field measurements, considering techniques such as infrared thermography and digital image correlation, together with data-driven methods and terms describing anomalous conditions, including damage, defects, and seepage.

When the number of retrieved studies was limited, the search strategy was adapted by relaxing some constraints, for example by using more general expressions or simplifying the combination of keyword groups.

3. AI for point-wise dam monitoring

In dam safety monitoring, anomaly detection is predominantly based on long-term point-wise measurements of the variables introduced in Section 2, with displacements and seepage-related quantities being the most commonly analysed.

Within this context, ML algorithms are primarily used to model the healthy-state behaviour of individual monitoring variables. Anomaly detection is then performed by analysing residuals, defined as the difference between measured responses and model predictions, and by evaluating their significance through statistical or model-based decision rules.

An illustrative example of single-variable behaviour modelling using a data-driven approach, together with the corresponding residual computation, is shown for two displacement variables of an Italian dam in Fig. 3.

Unlike vibration-based SHM, where ML is often applied directly for anomaly or damage pattern recognition, dam monitoring typically relies on it as a preliminary step preceding the actual diagnostic stage. More direct pattern-recognition strategies arise only when multiple monitoring variables are analysed jointly in a multivariate setting.

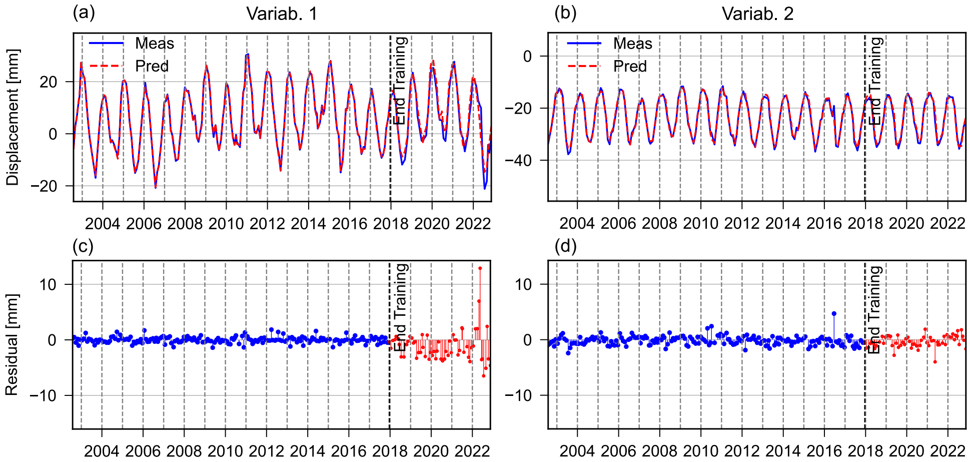

Fig. 3. Example of single-variable behaviour modelling and residual analysis for two displacement measurements. The upper panels (a-b) show measured and predicted responses, while the lower panels (c-d) report the corresponding residuals over training and testing periods.
Source: Adapted from [25].

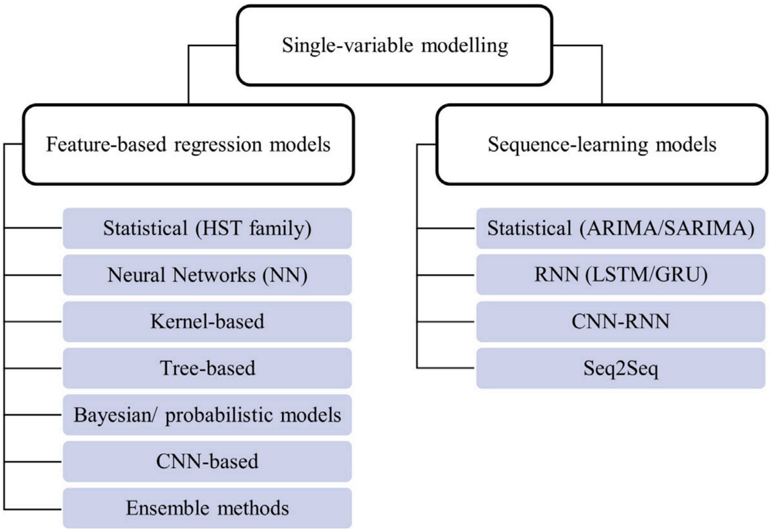

Fig. 4. Conceptual taxonomy of single-variable modelling strategies in dam safety monitoring, distinguishing between feature-based regression models and sequence-learning approaches.

The present section is structured to reflect this diagnostic progression, moving from single-variable behaviour modelling and threshold-based anomaly detection to multivariate analysis.

3.1. Single-variable diagnostic modelling

Single-variable modelling approaches can be broadly classified according to how temporal information is incorporated into the predictive formulation. Two distinct perspectives can be identified. The first considers regression models in which environmental and temporal effects are explicitly represented through engineered predictors derived from monitoring data, and the structural response is expressed as a function of these features. The second includes sequence-learning models, where monitoring records are treated directly as time series and temporal dependencies are learned implicitly from sequential data. Within each perspective, both ML and DL techniques are considered and further grouped into methodological families according to their underlying learning principles. This conceptual organization is illustrated schematically in Fig. 4.

Independently of the specific modelling approach, model implementation follows a common procedure structured into training and testing phases. Accordingly, monitoring data are partitioned into corresponding subsets: the training set is used for model fitting, while the testing set is reserved for the final performance assessment. A validation phase may also be included for model selection and hyperparameter tuning.

Model performance is primarily assessed on the testing set using standard error metrics, including the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE), its normalized form (NRMSE), and the coefficient of determination (R^2). Normalization is typically based on the range of the training data and allows comparability across variables.

3.1.1. Feature-based regression models

Data-driven modelling of dam behaviour has traditionally relied on regression-based approaches, in which the structural response is expressed as a function of environmental and operational variables. Early developments were based on statistical regression models, with the Hydrostatic-Seasonal-Time (HST) formulation representing the historical reference [26]. In this approach, dam displacement is decomposed into hydrostatic, seasonal, and long-term irreversible components through predefined functions, providing a simple and interpretable baseline.

Several extensions have been proposed to improve the representation of thermal effects. One group modifies the original formulation by introducing simplified heat-transfer-based representations, leading to models such as HTT (Hydrostatic–Temperature–Time) and related variants (HST-Grad, HST-Layer) [27–30]. Another group incorporates measured air temperature directly into the regression model, resulting in the HSTT (Hydrostatic–Seasonal–Time–Temperature) formulation, where daily temperature records enhance the empirical description of thermal loading [31–33]. More recent developments combine statistical regression with thermo-mechanical analyses by integrating Finite Element (FE) derived response components into the regression framework [34,35].

The HST model can be interpreted as a multilinear regression based on predefined predictors, while ML methods provide greater flexibility by learning nonlinear relationships directly from data. Among ML techniques, Neural Networks (NNs) were among the earliest approaches and are acknowledged in international guidelines [7]. Early applications relied on simple architectures such as Multilayer Perceptrons (MLPs) and Radial Basis Function Networks (RBFNs) [36,37]. More efficient formulations were later explored, including Extreme Learning Machines (ELMs) and their adaptive variants (e.g., AF–ELM), which address non-stationary behaviour through forgetting mechanisms [38]. Extensions incorporating Mean–Variance Estimation (MVE) also enable the direct estimation of prediction uncertainty [39,40].

Kernel-based methods, including Support Vector Regression (SVR), have been widely applied, showing good generalization when properly calibrated [41,42]. Variants such as Kernel ELM (KELM) and multi-output formulations have been proposed to improve predictive performance and enable the joint prediction of multiple monitoring points [43–45].

Tree-based ensemble models, particularly Random Forests (RF) and Gradient Boosted Regression Trees (GBRT), have gained increasing attention due to their robustness, ability to handle heterogeneous inputs, and limited preprocessing requirements. Comparative studies show a favourable balance between accuracy and complexity, with recent work highlighting differences among implementations such as XGBoost, LightGBM, and CatBoost [46,47]. These models also provide useful interpretability through feature importance and partial dependence analysis, and have been employed for data-driven threshold definition [48,49].

Within the probabilistic ML family, Gaussian Process Regression (GPR) and Relevance Vector Machines (RVM) provide both predictions and explicit uncertainty quantification. GPR models have been used to describe dam displacements as functions of hydraulic, thermal, and temporal effects, with particular attention to kernel selection [50], while an optimized RVM enabled sparse Bayesian solutions and achieved improved predictive accuracy compared with other methods [51].

More recently, DL models have been introduced. Convolutional Neural Networks (CNNs) have been applied as regression models by processing structured input features derived from HST variables [52]. In parallel, fully convolutional architectures have been proposed to predict spatiotemporal deformation fields from physically meaningful inputs such as hydrostatic pressure and temperature, providing prediction at both monitored and unmonitored locations [53]. In these cases, CNNs operate on structured spatial or feature representations rather than directly modelling temporal sequences.

From a quantitative perspective, the comparison results reported within the literature confirm the absence of a consistently superior ML model. In Babadi et al. [47] RF-based models achieve a MAE of 3 mm, slightly outperforming the compared NN model, which shows values of 3.8 mm. Similarly, Mata et al. [54] report only limited differences among SVR, NN, and RF, with MAE values ranging from 1.8 mm to 2 mm. In Alocén et al. [55], RMSE values range from 0.5 mm to 2 mm across different pendulum records from three dam case studies, again with no model consistently outperforming the others. Notably, among the time series analysed in that study, HST performs comparatively better for the two records characterized by strong temporal trends and limited data availability, suggesting a potential advantage under extrapolation conditions. Under standard conditions, ML models generally achieve better performance, although this advantage remains strongly dependent on calibration.

In particular, SVM and NN-based models are more sensitive to hyperparameter selection, whereas tree-based methods tend to be more robust to non-informative inputs and less sensitive to tuning. This has motivated the adoption of model-combination strategies, in which multiple model outputs are aggregated through weighted schemes to improve robustness and reduce extreme prediction errors [56].

Overall, comparison across case studies remains challenging, as results are often driven more by data-related choices, such as input definition, training period, and testing conditions, than by the modelling technique itself, with further variability arising from differences in signal characteristics (e.g., stationary or non-stationary behaviour).

Recent research has also expanded toward complementary aspects, particularly multi-output prediction and uncertainty-aware approaches. Multi-output strategies have been explored either through partition-based grouping of monitoring points or through CNN-based representations of spatial fields [45,53]. At the same time, increasing attention is being paid to uncertainty-aware formulations, capable of directly estimating prediction intervals.

Traditionally, uncertainty handling and early-warning definition in dam behaviour modelling have relied on residual-based thresholding applied to deterministic regression models, where uncertainty is inferred a posteriori from residual statistics. Common practices include thresholds based on residual standard deviation, empirical quantiles, or historical envelopes derived from training data [32,57,58].

Within this framework, several recent contributions have proposed adaptive residual-based strategies, which retain deterministic regression models but allow warning thresholds to evolve in response to changing conditions. For instance, one approach introduces adaptivity in the time domain by using moving statistics of the residual sequence computed over sliding windows [59], whereas another modulates threshold width according to local data density in the input space, combining regression models with density estimation techniques [60].

In parallel with these developments, beyond fully Bayesian approaches, a growing number of studies have extended deterministic regression models by directly incorporating prediction intervals or uncertainty bounds within the modelling phase. These include bootstrap-based methods, which propagate data uncertainty through resampling [61]; quantile regression techniques, which estimate conditional bounds of the response distribution without distributional assumptions [62]; and variance-aware learning schemes, in which dispersion is jointly estimated with the mean response during model training [40].

The ML and DL regression families discussed in this section, together with their corresponding methods, representative references, and, where applicable, associated warning strategies, are presented in Table 1.

3.1.2. Sequence-learning models

A complementary formulation of feature-based regression models considers dam monitoring records primarily as time series, focusing on their temporal structure. In this perspective, early modelling approaches originated from classical statistical methods. In particular, AutoRegressive Integrated Moving Average (ARIMA) models and their seasonal extension, Seasonal ARIMA (SARIMA), have been applied to dam displacement prediction, showing satisfactory performance in reproducing trends and recurring seasonal patterns typical of long-term monitoring data [65,66].

Building on these formulations, hybrid schemes combining regression-based components and SARIMA remain in use. Recent contributions include industrial-oriented MLR–SARIMA tools, in which a

Table 1
Overview of machine learning and deep learning regression models for dam behaviour prediction.

Model family	ML method	Output	References	Warning strategy
Neural networks	MLP, RBFN, ELM, hybrid HST–NN	Single output	[36–38,59,63]	Residual-based thresholds (moving statistics) [59]
Adaptive NN	AF-ELM, AF-OS-ELM-MVE	Single output	[39,40]	Mean Variance Estimation (MVE) [40]
Kernel-based	SVR, KELM, LSSVM	Single output	[41–44,61]	Bootstrap-based prediction intervals [61]
Multi-output kernel-based	Partition-based MSVM	Multiple outputs	[45]	–
Tree-based ensembles	RF, GBRT	Single output	[46–49,60,62,64]	Adaptive residual thresholds [60], Quantile regression [62]
Bayesian/probabilistic ML	GPR, RVM	Single output	[50,51]	Probabilistic prediction
CNN	CNN + Fully Connected NN	Single output	[52]	–
	Fully CNN	Continuous field	[53]	–
Ensemble/model combination	Stacking, greedy weighted stacking	Single output	[55,56]	–

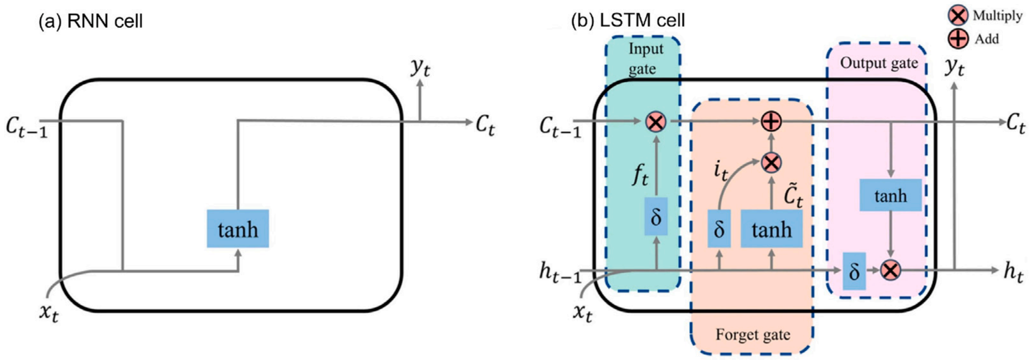

Fig. 5. Comparison between a standard RNN cell and an LSTM cell.
Source: Adapted from [69].

correlation analysis between displacement and reservoir level is used to guide the regression formulation, and the outputs of the regression and SARIMA models are combined through weighted aggregation [67]. Another study proposes SARIMA–Neural Network hybrid, where a NN is employed to correct systematic prediction errors of the SARIMA [68].

Beyond classical statistical formulations, DL models explicitly designed for sequential data have been increasingly applied to dam monitoring. Recurrent Neural Networks (RNNs) are among the most widely adopted for time-series prediction, as they propagate information across time steps through recurrent connections.

Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures represent structured extensions of the classical RNN. A schematic comparison between a standard RNN cell and an LSTM cell is shown in Fig. 5. Both LSTM and GRU introduce learnable gating mechanisms that regulate information flow, mitigating vanishing gradient effects and improving the modelling of temporal dependencies in non-stationary signals.

GRU models adopt a simplified gating structure with fewer parameters, which makes them attractive for large-scale or near-real-time monitoring applications [70]. LSTM networks, by contrast, incorporate an explicit memory cell that enhances the modelling of long-range dependencies and delayed environmental effects. Applications in dam monitoring include both deterministic and probabilistic implementations, as well as cases where training is supported by synthetic data generated from numerical models [71–73].

Other studies focus on input representation strategies, such as dimensionality reduction via Principal Component Analysis (PCA), compared with the more commonly adopted use of moving-average temperature indicators over multiple time windows [74]. More recently, bidirectional LSTM architectures have been explored to exploit information from both past and future temporal contexts within a sliding window [69].

Several approaches further combine recurrent architectures with convolutional feature extraction. In CNN–LSTM and CNN–GRU frameworks, convolutional layers are used to extract local temporal features from multivariate monitoring sequences before recurrent modelling [75,76]. In practice, convolution can be implemented either as a one-dimensional operation with the input variables treated as separate channels, or as a two-dimensional convolution in which the kernel spans the full feature dimension and slides only along the temporal axis. Both formulations are commonly used in the literature.

Several variations of these hybrid architectures have been developed. For example, some approaches extend the commonly used input factors by incorporating spatial dependencies through preprocessing steps such as clustering of monitoring points and decomposition of shared components, which are then used as additional inputs and weighted over time via attention mechanisms [77]. Other studies show that CNN–LSTM structures remain effective in the presence of complex damage-related phenomena, such as crack opening displacements, when appropriate feature screening is applied [78]. In another case, dual-stage architectures further refine the CNN–LSTM configuration by combining convolutional networks and LSTM units with a second recurrent stage trained on the residuals of the initial prediction to capture remaining stochastic components [79].

Another class of time-series models investigated in dam monitoring consists of encoder–decoder sequence to sequence (Seq2Seq) architectures equipped with attention mechanisms. In one approach, LSTM units form the encoder–decoder backbone and are combined with dual-stage attention to weight relevant hydrostatic, thermal, and time-related information [80]. In another, the Seq2Seq structure is built on a simplified recurrent unit, the Tiny Gated Unit (TGU), which reduces model complexity while retaining the ability to capture nonlinear temporal dependencies [81].

From a comparative perspective, sequence-learning models generally outperform traditional machine learning approaches, as they

Table 2
Overview of time-series–based machine learning and deep learning models for dam behaviour prediction.

Model family	ML method	Output	References	Warning strategy
Statistical time-series models	ARIMA, SARIMA	Single output	[66]	–
Hybrid SARIMA-based models	MLR–SARIMA, SARIMA–NN	Single output	[67,68]	–
Recurrent neural networks (GRU)	GRU	Single output	[70,82]	–
Recurrent neural networks (LSTM and variants)	LSTM, Bi-LSTM, Bayesian LSTM	Single output	[69,73]	Quantile regression, conformal calibration [69]
CNN–RNN hybrid architectures	CNN–LSTM, CNN–GRU	Single output	[75,76]	–
Extended CNN–RNN architectures	Attention-based CNN–LSTM, Dual-stage CNN–LSTM,	Single output	[77–79]	–
Seq2Seq with attention	LSTM–Seq2Seq, TGU–Seq2Seq	Single output	[80,81]	Quantile regression [81]

exploit both causal variables and the temporal evolution of the target signal. This trend is consistently observed across studies. For instance, in Hua et al. [75], where uplift pressure is predicted, MAE values decrease from about 0.83–0.89 m for ML models (SVM and NN) to values in the range of 0.15–0.23 m for deep learning models (CNN–GRU and GRU). Similarly, in Huang et al. [80], the error for a displacement variable reduces from about 1.04–1.17 mm to approximately 0.57 mm with advanced DL configurations.

LSTM and GRU models typically provide comparable performance, with no clear dominance of one over the other. However, hybrid architectures incorporating convolutional layers or attention mechanisms can yield further improvements. In Hua et al. LSTM and GRU achieve MAE values in the range of 0.23–0.33 m, while a CNN–GRU model reduces the error to about 0.15 m. In Huang et al. the transition from standard LSTM models (RMSE in the range of 0.80–0.92 mm) to Seq2Seq models leads to values around 0.55–0.57 mm.

Despite the inherent feature extraction capability of DL models, performance remains strongly dependent on the richness of the input data. In Cao et al. [77], a CNN–BiGRU–attention model achieves a MAE of about 0.30 mm when both load factors and spatially correlated displacements are included. The error increases to 0.77–0.95 mm when part of this information is removed, and up to 1.39 mm in purely autoregressive configurations. However, the inclusion of multiple heterogeneous inputs reduces transparency. Unlike traditional ML, where predefined variables (e.g., moving averages) allow a clearer interpretation of delayed effects, DL models learn these dependencies implicitly, making it more difficult to identify the driving factors behind the predictions.

Finally, also in this context, research has expanded toward complementary aspects, particularly the definition of prediction limits. These are typically based on quantile formulations, consistent with those used in regression-based settings (Section 3.1.1). For example, in the Seq2Seq-based study [81], the model is trained to predict conditional quantiles of the displacement response, directly yielding prediction intervals. Similarly, in the Bi-LSTM case [69], quantile regression is combined with conformal calibration, whereby the predicted bounds are adjusted using calibration residuals to improve interval reliability and achieve the desired coverage level.

An overview of the ML and DL sequence models reviewed in this section is provided in Table 2.

3.2. Multivariate anomaly detection

The limitations of single-variable anomaly detection are overcome by multivariate approaches that account for the joint behaviour of multiple monitoring variables. Early studies addressed this problem through supervised classification using labelled datasets generated from numerical simulations. Representative examples include the work of Mata et al. [28], who combined PCA and Linear Discriminant Analyses (LDA) on pseudo-experimental displacement data, and the subsequent study by Salazar et al. [83], who applied RF for multi-class damage classification based on scenarios generated by imposing displacements on abutments in FE simulations. While effective, these approaches are limited by their reliance on data representing predefined damage cases.

To reduce the need for labelled samples, unsupervised formulations trained on healthy-state observations have also been explored, for example through the application of one-class SVM to multivariate displacement vectors derived from simulated arch dam crack scenarios [84].

A different family of unsupervised approaches is represented by clustering-based strategies, which have been mainly investigated as tools for spatio-temporal pattern analysis rather than as standalone anomaly detection methods. Chen et al. [85] proposed a spatio-temporal fuzzy clustering method to identify zones with homogeneous deformation behaviour across different time periods. Building on this concept, Wang et al. [86] introduced a modified DBSCAN (Density-Based Spatial Clustering of Applications with Noise) formulation capable of identifying dense behavioural regions without requiring a predefined number of clusters. Lei et al. [87] further extended clustering toward anomaly interpretation by defining reference spatial deformation patterns and associated admissible envelopes, such that deviations combined with temporal anomaly detection can be interpreted as abnormal behaviour. Finally, Mata et al. [59] employed DBSCAN on multivariate residual features derived from prior NN models, explicitly using clustering for novelty detection. Overall, clustering-based methods in dam monitoring have been predominantly adopted as exploratory tools, with only limited extensions toward explicit anomaly detection.

More recently, DL has been used for multivariate anomaly detection through reconstruction-based and similarity-based formulations. In reconstruction-based approaches, Autoencoders (AEs) are trained under normal operating conditions to reproduce the multivariate structural response, with anomalies identified as deviations between observed and reconstructed data through thresholded reconstruction errors. However, their effectiveness depends not only on the model architecture, but also on how anomaly scores and thresholds are defined.

Nogara and Salazar [88] provide an example of an AE-based method in which each input consists of all monitored displacements at a given observation time. In this setting, the model implicitly learns the correlations among variables. Anomalies correspond to small, damage-induced structural variations, and detection is performed using variable-wise residuals rather than a global reconstruction error. This leads to improved sensitivity compared to the conventional approach and also enables damage localization. Nevertheless, the validation is entirely based on synthetic FE-generated data.

More complex architectures have been proposed, although they are often evaluated in simplified settings, where anomalies are introduced as point-wise or short-term offset variations rather than physically consistent damage scenarios. Shu et al. [89] present a Spatio-Temporal Variational Autoencoder (STVAE), a generative model combining recurrent structures and graph-based representations to capture temporal evolution while explicitly modelling relationships among variables through a graph structure. This allows both temporal and spatial anomalies to be detected. From a quantitative perspective, the model achieves very high precision (0.996), comparable to that of SVAE (Sequential VAE) and CNN-AE, while improving recall (0.48 compared with 0.24–0.36), leading to higher F1 scores. It also outperforms standard AEs and traditional statistical models such as ARIMA. However, these results are influenced by heterogeneous and model-specific

Table 3

Overview of anomaly detection approaches for dam monitoring, organized by learning paradigm.

Learning family	Reference	Method	Type of anomaly
ML – Supervised – Classification	Mata et al. [28]	PCA + LDA	Simulated structural damage (failure scenarios)
	Salazar et al. [84]	RF, SVM, One-class SVM (unsupervised)	Simulated structural damage (crack scenarios)
	Salazar et al. [83]	RF	Simulated structural damage (foundation/abutment slip)
ML – Unsupervised – Clustering	Chen et al. [85]	Spatio-temporal fuzzy clustering	Behavioural zoning
	Wang et al. [86]	Modified DBSCAN	Behavioural zoning
	Lei et al. [87]	Hierarchical clustering + envelopes	Measurement anomalies
	Mata et al. [59]	DBSCAN on residual features	Measurement anomalies
DL – Supervised – Contrastive learning	Wang et al. [90]	Spatio-temporal contrastive learning	Measurement anomalies
	Tutivén et al. [91]	Siamese neural network (binary classification)	Simulated structural damage (crack scenarios)
DL – Unsupervised – Reconstruction	Shu et al. [89]	Spatio-temporal variational autoencoder	Measurement anomalies
	Kang et al. [92]	MemAE–GAN	Measurement anomalies

threshold definitions, which are not standardized across methods, although the comparison remains indicative of the potential of this approach.

A further development is represented by the integration of adversarial learning within AE-based frameworks. In this direction, Kang et al. [92] develop a MemAE-GAN model, combining memory-augmented autoencoders with adversarial training, where a discriminator network is used to amplify discrepancies between real and reconstructed data. These approaches, also belonging to the broader class of generative models, aim to learn more expressive representations of normal behaviour. In the presented application, the model is applied at the level of individual measurement points and achieves very high performance. For one of the two analysed monitoring points, accuracy is 0.97 and F1 score is close to 0.98, slightly improving over AE-based variants and clearly outperforming classical ML methods. However, these results rely on synthetically generated anomalies together with Receiver Operating Characteristic (ROC)-based threshold calibration, which introduces an implicit level of supervision. Moreover, the performance is obtained in a point-wise setting, without exploiting multivariate dependencies.

A different strategy is adopted by similarity-based DL methods, which focus on on learning representations in a latent space rather than reconstructing the input. The model is trained on pairs of samples to map normal observations close to each other while pushing dissimilar ones farther apart, typically through a contrastive loss and shared encoding branches. After training, anomaly detection is performed based on distances in the learned embedding space or via a downstream classifier.

In dam monitoring applications, a spatio-temporal contrastive strategy has been proposed by Wang et al. [90], where multivariate time sequences are used to learn representations capturing both temporal evolution and inter-sensor relationships. The model is first pretrained in a self-supervised manner, exploiting unlabelled data, and then fine-tuned with a limited amount of labelled data, still requiring some supervision in the final stage. The method achieves accuracy and precision above 95% and F1 scores above 80% in most of the analysed measurement points.

Another study within this category formulates damage recognition as a supervised binary classification problem using a Siamese Neural Network (SNN) [91] and a dataset derived from a simulated cracked dam, where multivariate sequences are compared to learn a similarity function between normal and anomalous behaviour. In this case, the availability of labelled data is central to the training process, as pairs of samples must be explicitly constructed. Compared with ML classification methods such as RF and SVM, which show variable performance across scenarios, the SNN achieves more consistent results, with accuracy close to 100% in most cases and around 94% in the most challenging one, particularly improving generalization to unseen anomaly classes.

Overall, these works indicate that DL–based approaches can outperform classical supervised and unsupervised ML methods for anomaly detection. A structured overview of the reviewed contributions is provided in Table 3, where the methods are organized according to their learning strategy.

3.3. Advances and open challenges

The literature on AI-based point-wise dam monitoring shows a clear methodological evolution from static input–output relationships toward the learning of complex temporal patterns, driven by recent advances in DL. This transition is also associated with a shift from single-output to multi-output formulations, enabling a more comprehensive representation of the structural response. More recently, generative AI approaches have introduced a further step, with Large Language Models (LLMs) enabling flexible representations of time series by treating them as structured sequences and integrating multisource monitoring data [93].

A similar transition can be observed in anomaly detection. The field has progressed from residual-based statistical thresholding applied to single predictive models toward multivariate pattern recognition. This includes supervised approaches, such as ML classification models, as well as representation-learning techniques and unsupervised methods, such as clustering, autoencoder and GAN-based approaches within the broader framework of generative AI. While these methods can capture complex patterns, the detected anomalies are often difficult to interpret, especially when they are not explicitly represented in the training data. In this context, emerging LLM-based methods suggest a further shift toward anomaly interpretation, supporting explanation and decision-making processes [94].

Despite these developments, the main limitation remains the availability of real monitoring data. Public datasets are scarce due to confidentiality constraints. Initiatives such as the 16th ICOLD Benchmark Workshop (BW) provide a limited number of real time series for comparing displacement and leakage prediction methods [95]. Moreover, these datasets mainly represent normal operating conditions, while real anomalies or damage cases are rarely available. As a result, validation often relies on simulated scenarios.

From an application perspective, the integration of ML approaches into operational monitoring remains limited. Traditional models, such as HST, are still widely adopted due to their robustness and interpretability. This tendency is also evident in industrial contributions within the cited ICOLD BW. For instance, Simon [32] proposes an extended HST-based formulation, including a temperature-related component for displacement modelling, while NNs are applied only for leakage prediction. Similarly, Corigliano et al. [67] employ ensemble approaches combining linear regression and SARIMA models, indicating that simpler, interpretable models are still preferred.

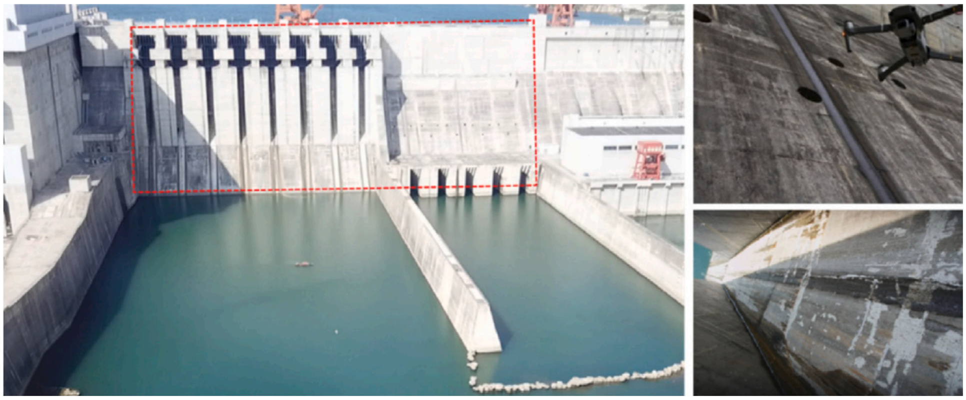

Fig. 6. UAV-based visual inspection of a concrete dam in the Jialing River Basin.
Source: Reproduced from [98].

To address these limitations, two complementary research directions are emerging, aiming to improve both the reliability of data-driven models under extrapolation and their practical usability within engineering workflows.

The first direction focuses on hybrid approaches combining FE simulations with ML. These strategies improve predictive performance, particularly under conditions not represented in the training data. More specifically, this integration can occur in three ways: (i) by directly using FE predictions as input features for ML models, as in the work of Pirker and Zenz [71], or by incorporating FE-derived terms within extensions of the HST formulation [35]; (ii) by employing ML methods to learn and correct the residuals between FE predictions and observed data; and (iii) by generating synthetic input–output pairs through FE simulations and using them to train ML models under a wider range of operating conditions [96]. While most applications focus on predictive modelling, similar hybridization concepts could also be applied to anomaly detection.

The second, more recent direction concerns LLM-based systems for orchestrating SHM workflows. LLMs can serve as high-level interfaces capable of coordinating multiple components, including prediction models, anomaly detection algorithms, and data management tools [94]. These systems rely on embedding-based representations to integrate heterogeneous data, retrieval-augmented mechanisms to incorporate prior knowledge, and agent-based or tool-calling strategies to activate specialized models. At present, the limited number of applications mainly integrates data-driven components. Extending these frameworks to include FE simulations represents a natural next step, enabling a tighter coupling between numerical modelling, monitoring data, and domain knowledge within a unified analysis system.

4. AI for vision-based dam inspection

Visual inspection is a key component of dam safety assessment, as many deterioration mechanisms manifest through surface-visible features such as cracks, spalling, or material degradation. Traditional inspections are increasingly complemented by image-based surveys using Unmanned Aerial Vehicles (UAVs) for exposed surfaces and Remotely Operated Vehicles (ROVs) for submerged zones, enabling safer and more extensive data acquisition. As illustrated in Fig. 6, such surveys provide both wide coverage and detailed visual assessment through the acquisition of RGB images using optical cameras. In some cases, UAVs or ROVs are complemented by range-based sensing technologies, such as time-of-flight cameras and LiDAR (Light Detection and Ranging) systems, or sonar sensors for underwater environments. The collected data, including both visual and range information, can support automated diagnostic analysis and spatial localization of damage, particularly when three-dimensional reconstruction is involved [97].

Image-based information is naturally addressed within the domain of computer vision, where damage is inferred from surface appearance rather than from structural response. General reviews on DL for image processing provide the methodological background for these applications, outlining the main vision tasks relevant to inspection, including image enhancement, classification, object detection, and semantic segmentation [99–101]. Within the broader field of civil infrastructure inspection, a wide range of image-based damage-identification problems has been investigated. As a representative example, crack recognition has received significant attention, with both ML and DL approaches explored [21,102,103]. Focusing exclusively on DL-based methods, the following subsections review methods developed for dam applications, organized by objective: (i) crack inspection and quantification; (ii) multi-damage identification, addressing not only cracks but also other surface deterioration phenomena; and (iii) the same inspection objectives applied to underwater surfaces, which are discussed separately due to their distinct environmental and data-related challenges. This classification, together with the methodological approaches discussed in the following sections, is schematically summarized in Fig. 7.

Since these tasks differ in their formulation, their performance is reported across different studies using task-dependent metrics. In image classification, accuracy, precision, recall, and F1-score are commonly adopted to quantify the ability to correctly identify a given class (e.g., the presence of cracks) at the image level.

In semantic segmentation, the evaluation is performed at the pixel level, where each pixel can be interpreted as an individual classification unit. In this context, precision and recall can still be defined, although overlap-based metrics are more frequently used, in particular the Intersection over Union (IoU), which quantifies the agreement between predicted and reference regions. When multiple classes are considered, IoU is typically averaged across classes, leading to the mean Intersection over Union (mIoU).

For object detection, performance is generally assessed using mean Average Precision (mAP), which summarizes the precision–recall trade-off. A detection is considered correct only if the predicted bounding box sufficiently overlaps with the ground truth, as quantified by IoU, which therefore serves as the criterion for defining true positives in the computation of mAP.

In the following, reported improvements in accuracy refer to these task-specific evaluation measures.

4.1. Crack detection and quantification

Crack inspection in concrete dams using DL is currently dominated by semantic segmentation models, which enable pixel-level recognition of cracks in images. In this formulation, cracks are manually delineated in the training dataset by assigning class labels to individual pixels, allowing the trained model to generate crack masks on new images.

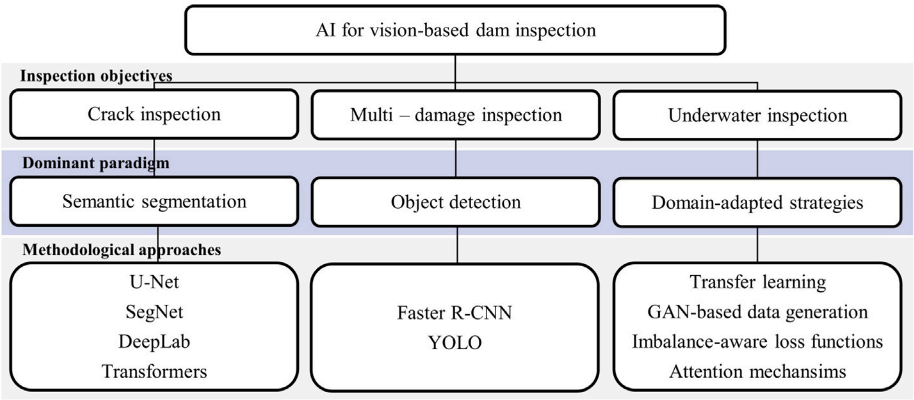

Fig. 7. Taxonomy of AI approaches for vision-based dam inspection, organized by inspection objective, dominant paradigm, and methodological implementation.

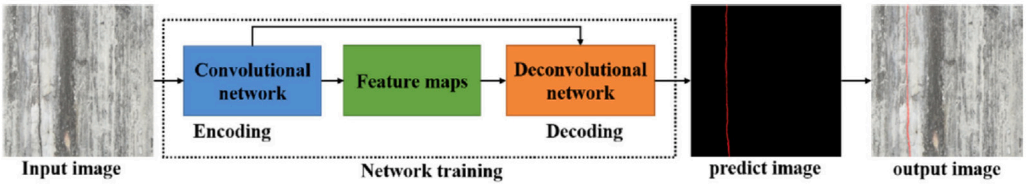

Fig. 8. Architecture of the DL-based crack detection network, illustrating a fully convolutional encoder–decoder structure for pixel-level crack segmentation. *Source:* Reproduced from [98].

These masks can then support geometric quantification of crack length, area, or width through conventional image-processing algorithms.

Most studies rely on fully convolutional encoder–decoder architectures, in which an encoder extracts hierarchical features and a decoder restores spatial resolution to produce dense crack masks, as illustrated in Fig. 8.

Widely adopted models include U-Net and SegNet, which preserve fine spatial details through skip connections and pooling indices, respectively. Building on these architectures, several specific enhancements have been proposed, mainly introducing attention mechanisms or lightweight modifications to improve feature extraction and deployability.

Feng et al. [98] trained a modified SegNet with additional auxiliary loss branches on UAV crack images from a single dam, achieving 80.45% recall and 80.31% precision. Performance was only marginally better than standard U-Net and SegNet baselines, while larger improvements were observed relative to simpler Fully Convolutional Networks (FCNs), which reached 71.53% recall and 72.57% precision. Subsequently, Chen et al. [104], using the same inspection images but a different dataset preparation strategy, reported substantially improved results with an attention-enhanced U-Net (90.81% precision and 81.54% recall), outperforming both the previous Feng model (85.49% precision, 37.49% recall) and the FCN baseline (79.65% precision, 65.34% recall), confirming the benefit of attention mechanisms under complex surface textures.

DeepLab-based models adopt a different strategy, using atrous convolutions and Atrous Spatial Pyramid Pooling (ASPP) modules to enlarge the receptive field and aggregate multi-scale contextual information, as illustrated in Fig. 9. Using real crack images from a dam inspection case study, Wu et al. [105] achieved 83.23% mIoU and 66.96% crack IoU with a lightweight MobileNetV2–DeepLabV3+, outperforming the U-Net benchmark tested in the same study (79.16% mIoU and 58.87% crack IoU).

Other recent works incorporate transformer-based components to improve global context modelling under challenging imaging conditions such as low contrast or motion blur. Moreover, when combined with photogrammetric reconstruction, segmentation outputs can be mapped onto three-dimensional dam surface models for 3D crack visualization [106].

Despite these algorithmic improvements, the main limitation remains data availability, as most studies rely on proprietary datasets provided by individual dam owners and related to single case studies, typically involving only a limited number of annotated crack patches. This also raises concerns regarding domain shift when models are transferred to new dams without specific retraining.

To investigate possible mitigation strategies, Xu et al. [107] combined public crack datasets from other domains, including open-source road and concrete surface crack images, with synthetic dam-specific images generated through virtual 3D rendering and image-based crack insertion onto real dam textures. The resulting hybrid dataset was used within a two-stage object-detection and segmentation framework and evaluated on real crack images acquired from the target dam. Incorporating the generated dam-specific images improved AP from 24.7% when training on generic crack data alone to 53.9%, highlighting the importance of domain-adapted augmentation even when synthetic rather than real dam images are used for training.

For crack detection, image classification using CNNs remains relevant as a simpler alternative for first-level screening, where the task is limited to determining whether a crack is present in an image or patch. In this setting, data scarcity is commonly addressed through transfer learning from large-scale datasets such as ImageNet. Using an ImageNet-pretrained EfficientNetB0 fine-tuned on borehole inspection imagery from a dam, Khan et al. [108] achieved 91% crack/non-crack classification accuracy, with compact EfficientNet variants outperforming several more complex CNN architectures, although all tested models achieved accuracies above 84.

To improve deployability, Zhang and Bao [109] combined transfer learning and knowledge distillation to obtain a lightweight ResNet18 model, pretrained on mini-ImageNet and further distilled on a generic concrete crack dataset rather than on dam-specific crack images. The resulting model achieved 98.39% classification accuracy on UAV images of a dam spillway, suggesting that staged transfer-learning strategies can support compact deployable models and partially compensate for the scarcity of dam crack datasets.

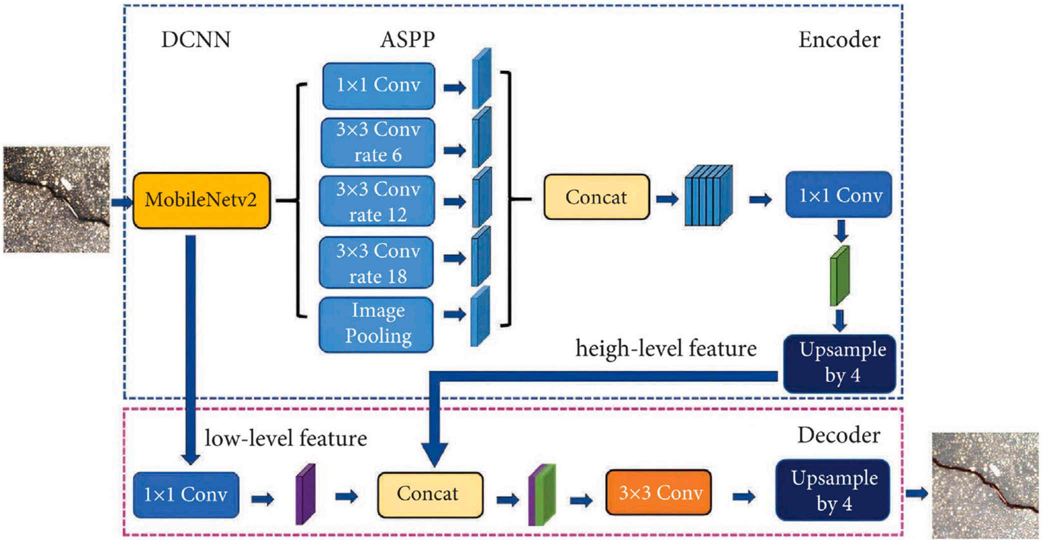

Fig. 9. Schematic representation of a DeepLab-based architecture for crack segmentation, illustrating the use of dilated convolutions and Atrous Spatial Pyramid Pooling (ASPP) modules.
Source: Reproduced from [105].

Finally, object detection approaches can also be adopted for rapid crack identification, where cracks are localized through bounding boxes rather than delineated pixel-wise [110]. However, these methods are generally more suitable for rapid localization of surface damage than for detailed geometric crack analysis, as discussed in the following section.

4.2. Multi-damage identification

Beyond crack-only inspection, several vision-based studies address multi-damage detection, where different types of surface deterioration are identified simultaneously from images. In this setting, the problem formulation often shifts from pixel-level segmentation to object detection, in which each damage instance is represented by a class label and a bounding box. This formulation is well suited to the heterogeneous nature of dam deterioration (e.g., cracks, spalling, efflorescence, exposed reinforcement, pits, or erosion), for which pixel-wise annotation would require substantially greater labelling effort.

Object detection methods are commonly classified into two-stage and single-stage detectors. Two-stage detectors, such as Faster R-CNN, first generate candidate object regions and subsequently refine their classification and bounding-box localization, generally providing higher localization accuracy at the expense of computational efficiency. In dam inspection, Huang et al. [111] improved Faster R-CNN through ImageNet pretraining, dam-specific anchor configurations tailored to the characteristic shapes of different defect classes, and data augmentation via cropping and flipping, thereby increasing viewpoint and scale variability in the training set. Trained on images acquired from multiple dams under non-uniform illumination and viewing conditions, the model achieved an overall mAP of 0.887, improving upon baseline Faster R-CNN performance (0.854), with the most significant gain observed for spalling detection, whose AP increased from 0.888 to 0.948.

Single-stage detectors, by contrast, remove the region proposal stage and directly predict bounding boxes and class probabilities in a single forward pass, followed by post-processing steps such as confidence thresholding and non-maximum suppression. Among these, YOLO-based models are increasingly preferred for UAV inspections due to their superior inference speed. Recent studies show a trend toward enhanced and lightweight YOLO variants, consistently accompanied by dedicated data augmentation strategies. Li and Bao [112] proposed a compressed YOLOv5-based detector combining adaptive feature fusion, pruning, and knowledge distillation, and trained it using geometric augmentations (cropping and flipping) together with photometric perturbations of colour and intensity to improve robustness to variable imaging conditions. The resulting model achieved mAP = 0.894, substantially outperforming both Faster R-CNN (0.547) and YOLOv4 (0.675) on the same dataset.

Zhao et al. [113] enhanced YOLOv5s with transformer-based modules together with random flipping and mosaic data augmentation, in which four training images are merged into a composite sample to increase scale diversity and contextual variability. The model was trained and tested on images collected from multiple concrete dam surfaces. The detector achieved 79.8% mAP, improving baseline YOLOv5s performance by 3.8 percentage points, while enabling centimetre-level 3D defect localization through integration with photogrammetric reconstruction.

Only limited work has addressed multi-damage detection through semantic segmentation. A notable example is Hong et al. [114], who combined multi-class semantic segmentation with 3D reconstruction for spillway inspection. Their training set comprised images collected from a dam using a wall-climbing robot and was expanded through rotation, flipping, brightness adjustment, and Gaussian blur augmentation to improve invariance to orientation, illumination, and image quality degradation. The dataset was further enriched with publicly available crack and spalling images from external concrete defect datasets. The resulting framework improved baseline segmentation performance by approximately 8 percentage points, reaching 0.61 mIoU, and reduced defect counting error from 9.5 to 1.75 instances through voxel-based fusion of overlapping segmented views.

Overall, multi-damage inspection prioritizes damage classification and localization over precise geometric measurement, making object detection the dominant paradigm. While two-stage detectors remain competitive when localization accuracy is critical, YOLO-based one-stage approaches are increasingly favoured for practical UAV deployment due to their real-time capability. Across the reviewed studies, augmentation and transfer-learning strategies are recurrently adopted alongside architectural refinements, reflecting the persistent challenge posed by the limited availability of dam-specific annotated datasets.

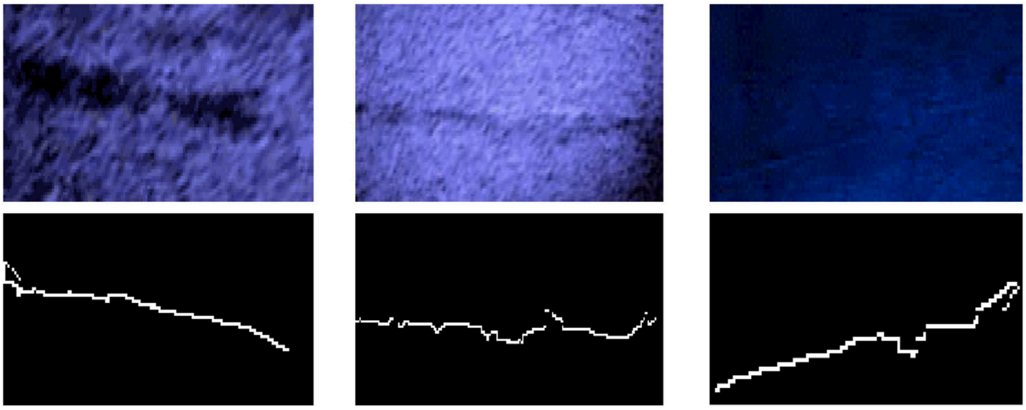

Fig. 10. Example of a sonar-based crack detection process for underwater dam inspection. The upper row shows representative input sonar images, while the lower row illustrates the corresponding segmentation outputs highlighting detected crack paths.
Source: Adapted from [116].

4.3. Underwater inspection methods

Underwater inspection of concrete dams presents substantially greater challenges than exposed-surface surveys due to poor visibility, colour distortion, non-uniform illumination, motion blur, and frequent occlusions from the ROV platform itself. In addition, the limited availability of annotated underwater inspection data and the strong imbalance between damaged and undamaged samples hinder the direct application of conventional vision models. Consequently, the literature places particular emphasis on improving robustness to degraded imaging conditions and on mitigating data scarcity through transfer learning, augmentation, and synthetic data generation.

For rapid screening applications, image classification has been explored using architectures specifically adapted to underwater imagery. Zhu et al. [115] enhanced a VanillaNet-based CNN with imbalance-aware loss weighting, label smoothing, and frequency-domain feature enhancement modules aimed at improving crack feature extraction under degraded underwater imaging conditions. The model classified six underwater crack morphologies and achieved 90.75% average accuracy, compared with 87.94% for ResNet. Training was performed on a limited dataset derived from only 250 original underwater dam images extracted from surveillance videos. This dataset was expanded through augmentation and further enriched with non-underwater dam crack images due to the scarcity of underwater samples. However, validation was restricted to internal data splits, leaving the robustness of the model across unseen underwater inspection environments and different dam conditions unverified.

When optical visibility becomes insufficient, alternative sensing modalities may be required. Shi et al. [116] addressed crack detection using dual-frequency sonar images rather than RGB imagery, demonstrating that acoustic imaging can enable crack identification in visually inaccessible conditions, as illustrated in Fig. 10. Nevertheless, the approach was validated on only 20 sonar images from a single dam and relied on handcrafted processing techniques, limiting conclusions regarding general applicability.

For pixel-wise crack delineation, Li et al. [117] proposed a lightweight LinkNet-based segmentation model combining focal loss, which assigns greater weight to difficult and underrepresented crack pixels to mitigate crack/background imbalance, with a staged transfer-learning strategy. The network was progressively adapted from ImageNet to a structural crack dataset comprising 14,268 images and subsequently fine-tuned on 855 annotated underwater crack images. The final model achieved 0.892 mIoU, outperforming U-Net (0.788) and DeepLabv3+ (0.787). Different fine-tuning configurations were also investigated, with the best performance obtained by freezing only the first encoder block during underwater-domain adaptation. However, validation again remained limited to splits derived from the same original dataset, and the actual transferability of the learned representations across independent inspection campaigns was not assessed.

Beyond transfer learning, synthetic data generation has emerged as a promising strategy to reduce dependence on scarce underwater datasets. Huang et al. [118] employed a CycleGAN-based image translation framework to convert above-water crack images into synthetic underwater-like counterparts. The model was trained using 984 above-water crack samples together with 678 unpaired underwater ROV images to learn the target visual domain. The generated synthetic data were subsequently used to progressively replace the real underwater subset of the training set. Above-water crack images were retained in all configurations, and replacement ratios from 25% to 100% were investigated. Results showed only limited performance variation even when synthetic data fully substituted real underwater training data. Specifically, classification accuracy remained between 95.6% and 100% for the evaluated ResNet models, object detection mAP decreased only from 82.59% to 79.68% with YOLOv5, and segmentation mIoU increased from 77.68% to 82.45% with U-Net. These findings suggest that synthetic generation can effectively support model training when real underwater images are unavailable or limited.

Recent work has also extended underwater inspection toward multi-damage detection. Kang et al. [119] formulated the task as object detection of cracks, spalling, and exposed reinforcement using a YOLO-based detector enhanced with deformable convolution networks, coordinate attention mechanisms, and an improved loss function. Trained on 2194 underwater images collected from multiple dams, and supplemented with additional samples generated through mosaic augmentation, the model achieved 84.5% mAP, improving baseline YOLOv8n performance by 3.2 percentage points. The framework was implemented on an ROV-mounted processing system and demonstrated in a pool experiment with artificially reproduced damage, illustrating the feasibility of real-time onboard damage screening. However, no quantitative performance results were reported for this experimental validation.

Overall, underwater dam inspection mirrors surface inspection in employing classification, segmentation, and object-detection paradigms, but places stronger emphasis on data-centric strategies to compensate for the scarcity and degraded quality of underwater imagery. Although reported performance is generally promising, evaluation remains predominantly confined to internal train-test splits derived from the same original datasets. Consequently, the true generalization capability of current models remains difficult to assess.

4.4. Advances and open challenges

Based on the review of the three main vision-based inspection tasks considered, whose corresponding studies are summarized in Table

Table 4
Overview of vision-based inspection methods for damage detection in concrete dams.

Reference	Environment	Sensor/Data	Target damage(s)	Formulation	Core method/Model family	3D info
Feng et al. [98]	Surface	UAV RGB	Crack	Semantic segmentation	Modified SegNet	No
Chen et al. [104]	Surface	UAV RGB	Crack	Semantic segmentation	Modified Attention U-Net	No
Wu et al. [105]	Surface	RGB	Crack	Semantic segmentation	DeepLabV3 with MobileNet	No
Zhao et al. [106]	Surface	UAV RGB	Crack	Semantic segmentation	Multi-Path Vision Transformer	Photogr.
Xu et al. [107]	Surface	UAV RGB + synthetic	Crack	Object detection	YOLOX	No
Khan et al. [108]	Gallery	RGB	Crack	Classification	CNNs + transfer learning (EfficientNet, etc.)	No
Zhang and Bao [109]	Surface	UAV RGB	Crack	Classification	ResNet + knowledge distillation	No
Li et al. [110]	Surface	UAV RGB	Crack	Object detection	YOLOv8	No
Huang et al. [111]	Surface	RGB	Multi-damage	Object detection	Faster R-CNN	No
Li and Bao [112]	Surface	UAV RGB	Multi-damage	Object detection	Improved YOLOv5 + knowledge distillation	No
Zhao et al. [113]	Surface	UAV RGB	Multi-damage	Object detection	YOLOv5s-HSC	Photogr.
Hong et al. [114]	Surface	Robot RGB	Multiple defects	Semantic segmentation	Modified U-Net	Simultaneous loc. and mapping
Zhu et al. [115]	Underwater	ROV RGB	Crack	Classification	Lightweight CNN	No
Shi et al. [116]	Underwater	Sonar	Crack	Rule-based	Clustering + tensor voting	No
Li et al. [117]	Underwater	ROV RGB	Crack	Semantic segmentation	LinkNet + transfer learning	No
Huang et al. [118]	Underwater	RGB	Crack	Image generation + classification/object detection/semantic segmentation	CycleGAN + CNNs/Faster R-CNN, YOLOX/DeepLabV3, HRNet, U-Net)	No
Kang et al. [119]	Underwater	ROV RGB	Multi-damage	Object detection	Modified YOLOv8	No

4, the assessment of concrete dams appears to be evolving along two complementary methodological directions. On one side, recent approaches increasingly adopt specialized architectural and algorithmic refinements, including attention mechanisms, transformer-based modules, and modified loss functions to better address class imbalance. On the other, a parallel line of development focuses on enhancing the training pipeline through dedicated data-related strategies.

This second trend largely reflects the need to compensate for the limited availability of annotated dam-specific image datasets for both exposed and underwater inspections. The reviewed studies propose a broad range of solutions, including conventional augmentation of case-specific inspection images, transfer learning from generic or structurally related datasets, enrichment through external public crack and defect repositories, and synthetic image generation via rendering, image-translation techniques, and more recently, generative AI methods.

In particular, GANs are used to produce realistic defect images and expand training datasets, while more recent diffusion models enable the creation of high-fidelity images [120], even from structured or textual inputs. At the same time, generative models are employed for image restoration and enhancement under challenging acquisition conditions, including low-light, noise, and underwater environments. These approaches show significant potential to improve both data quantity and quality. However, the lack of sufficiently large and diverse real-world dam datasets remains a critical limitation, as it restricts the reliable evaluation and validation of models developed using synthetic data.

A first practical challenge, therefore, concerns model generalization. Despite the promising accuracies reported, the real-world robustness of current approaches remains insufficiently demonstrated. Most studies rely on train/test splits derived from the same dataset, typically collected from a single dam or a limited group of dams within the same inspection campaign or acquisition setup, which likely leads to optimistic performance estimates relative to deployment on new case studies. In practice, application to a different structure may introduce substantial domain shifts due to variations in geometry, surface condition, environmental exposure, and imaging setup. Although a limited number of recent works begin to explore more heterogeneous testing scenarios, systematic validation across independent dams or inspection campaigns remains largely absent. More broadly, the field still lacks publicly available benchmark datasets and unified damage taxonomies, which prevent objective comparison across studies and the definition of shared implementation standards.

Another practical issue is the deployment of the methods in operational monitoring environments. Accordingly, recent research has increasingly focused on related requirements and operational constraints beyond simply improving detection accuracy. Some studies explicitly focus on lightweight or real-time implementations suitable for onboard UAVs/ROVs, while others integrate defect detection with photogrammetric reconstruction to locate damage on three-dimensional dam models, thereby enabling more structured inspections and digital-twin-oriented applications.

Beyond these methodological developments and challenges, recent advances in multimodal AI have also opened new possibilities for inspection data. In particular, LLM-based frameworks support zero-shot detection, i.e., the ability to identify damage patterns without task-specific training, as well as automated report generation and human–machine interaction [121]. While these methods show promising performance in structured tasks, their accuracy generally remains lower than that of specialized models. For this reason, their role is currently more aligned with interpretation and decision support than with primary damage detection.

5. AI for non-destructive testing in dam monitoring

Non-destructive testing (NDT) techniques are an important complement to dam surveillance, alongside continuous monitoring based on point sensors and vision-based inspection. Since dams are primarily subjected to environmental loads, these natural actions can be exploited as excitation sources, and the resulting structural response can be measured and interpreted with the support of physics-based models. In this context, the acquisition of quantitative data is essential. Non-contact optical methods provide surface measurements and are particularly well suited to large concrete structures, as they are compatible with both ground-based and UAV-based inspections. In particular, Infrared Thermography (IRT) and Digital Image Correlation (DIC) are surface-based techniques capable of providing spatially dense, full-field measurements of temperature and displacement or strain fields, respectively, over the investigated areas.

A distinctive feature of IRT and DIC lies in the dual nature of the data they produce. When treated as images, thermographic maps and

DIC outputs can be analysed using computer-vision and DL techniques for damage identification and anomaly detection, such as the localization of seepage zones, cracks, or other surface-related irregularities. At the same time, these measurements represent continuous physical fields that extend beyond purely image-based interpretation.

They can support FE-assisted analysis and model parameter updating by providing spatially distributed quantitative information that is rarely accessible through point-wise instrumentation alone. AI can act as a surrogate for computationally intensive FE models, enabling more efficient inverse analyses and facilitating the calibration and continuous updating of digital twins of the considered structures.

However, although optical techniques, AI-based image analysis, and FE-based model updating have each been investigated in dam engineering, their integrated application remains limited. Accordingly, the following sections do not aim to provide an exhaustive review, but rather discuss representative studies addressing these aspects individually, which may support future integration for specific monitoring and modelling tasks.

5.1. Infrared thermography

IRT provides full-field measurements of surface temperature, contributing to dam assessment in two distinct ways: as a diagnostic imaging modality for identifying thermally detectable damage and leakage anomalies, and as a source of distributed thermal information for predictive and physics-based modelling.

The use of thermography in dam inspection is well established, particularly for detecting water leakage and seepage zones on downstream faces. Early applications rely primarily on qualitative interpretation of thermal patterns supported by expert judgement. Henriques and Ramos [122], for example, applied thermographic surveys to three Portuguese dams to monitor surface pathologies, highlighting cold anomalies associated with active seepage and comparing thermal observations with conventional visual inspections. In this context, thermography acts mainly as a rapid screening tool, enabling significantly faster and broader inspection coverage than manual surveys.

A further development concerns the automation of thermographic interpretation. From a data-analysis perspective, thermographic measurements can naturally be treated as images, making it possible to apply DL-based computer vision techniques analogous to those used for RGB imagery. This broader potential is discussed by He et al. [123], who review DL applications in infrared machine vision across several contexts, including SHM.

Within dam-related applications, automated thermogram interpretation has been formulated using both semantic segmentation and object-detection paradigms. The former enables pixel-wise delineation of thermal anomalies. In this context, Pozzer et al. [124] proposed a multiclass CNN-based method trained on thermal images acquired from a highly deteriorated buttress dam to identify delamination, cracks, spalling, and repair patches. Their results demonstrated the feasibility of automated pixel-wise defect mapping from raw thermograms, including the detection of subsurface deterioration not readily visible in RGB imagery, while also highlighting the need for larger and more diverse thermal datasets to improve generalization.

Similarly, Wang et al. [125] employed a segmentation-based formulation for seepage delineation from UAV-acquired thermograms of an embankment dam. They proposed a modified U-Net architecture incorporating temperature-based prior masks generated by identifying the pixels whose temperatures matched field-measured seepage temperatures. By guiding the network toward thermally compatible regions, the method reduced false detections caused by vegetation and other interferences, although its reliance on site-specific thermal priors may limit transferability across different dams and environmental conditions.

Alternatively, object-detection approaches provide coarser but computationally lighter localization when approximate anomaly positioning is sufficient. Zhao et al. [126] demonstrated the applicability of an enhanced YOLO model for UAV dam inspection in the context of pedestrian detection, highlighting the potential transferability to leakage recognition. This concept was subsequently implemented by Wang et al. [127], who extended the object-detection formulation to the joint identification of seepage and surface personnel.

Alongside these developments, other studies have explored DL-based preprocessing strategies. For example, Zhu et al. [128] investigated GAN-based super-resolution techniques to enhance infrared thermograms for leakage inspection, suggesting that image-enhancement methods may improve the interpretability of low-resolution or long-distance thermal acquisitions prior to automated defect analysis.

The studies discussed above mainly consider single thermographic surveys. Repeated thermographic acquisitions may improve diagnostic interpretation by introducing a temporal dimension to the analysis. In this context, recent infrared NDT research has begun exploring spatio-temporal DL methods that exploit thermal image sequences rather than isolated thermograms. For example, Saeed et al. [129] combined automated defect localization with NN analysis of pixel-wise thermal contrast curves to estimate defect depth from pulsed thermography sequences, while Luo et al. [130] proposed hybrid spatial–temporal architectures integrating recurrent LSTM modules within segmentation networks to improve defect delineation in low-contrast thermographic sequences. Although developed for active thermography in composite-material inspection, these studies highlight the potential of exploiting temporal thermal evolution as an additional discriminative feature. In dam applications, as seepage- or damaged-affected regions may exhibit distinct heating and cooling behaviour, monitoring the temporal evolution of thermal anomalies with successive inspections (over daily cycles or annual periods) may improve their characterization and help distinguish persistent structural issues from transient environmental effects. However, the uncontrolled nature of passive thermal excitation makes such temporal modelling substantially more challenging than in active thermography.

Beyond its diagnostic role, thermography can also support predictive and physics-based modelling by providing spatially distributed thermal fields. This information may improve predictive accuracy when thermal loading constitutes a dominant component of dam response, as illustrated in Fig. 11 for an Italian dam. In this context, thermographic measurements can provide distributed thermal boundary information not accessible through conventional point-wise instrumentation. For example, Yang et al. [131] proposed a Physics-Informed Neural Network (PINN) that combine sparse sensor measurements with spatial thermal observations while enforcing heat-transfer consistency through physically derived loss terms, thereby providing more physically coherent reconstruction of dam temperature fields.

5.2. Digital image correlation

DIC is a non-contact, full-field measurement technique widely used in the SHM of civil infrastructure to measure surface displacement and strain fields. By tracking surface texture or speckle patterns through image correlation, DIC computes displacement fields from the relative motion of image subsets, resulting in spatially distributed deformation measurements that complement traditional point-wise sensors. Its methodological foundations and broad applicability in SHM are well documented, primarily in bridges and other civil engineering structures [132].

Applications of DIC to dam monitoring remain relatively limited. Existing studies mainly address specific experimental or operational scenarios. Lei et al. [133] applied three-dimensional DIC to a physical model of an arch dam subjected to impact loading, demonstrating that full-field strain measurements can identify localized strain concentrations and anticipate crack initiation prior to the appearance of visible damage. While this study highlights the potential of DIC for early damage detection in dam structures, its validation is restricted to laboratory-scale conditions.

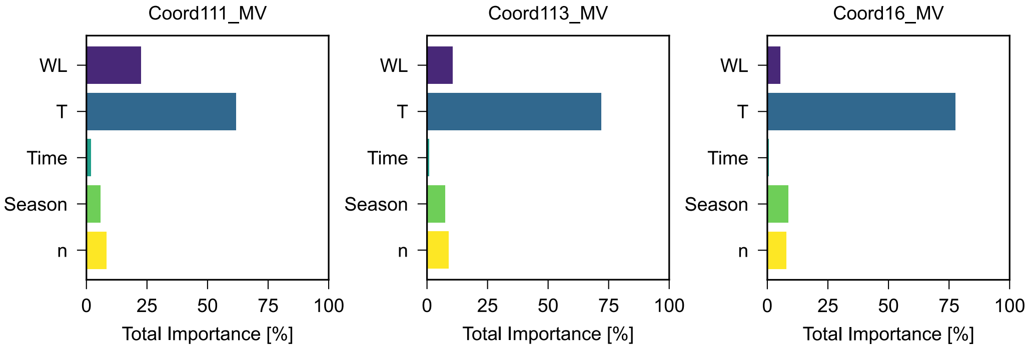

Fig. 11. Relative importance of different environmental and temporal effects at four representative dam response points for two Italian case studies, highlighting the dominant influence of temperature (T).
Source: Adapted from [25].

Field applications on real dams are rarer and are mostly focused on displacement monitoring rather than direct damage detection. Zaczek-Peplińska et al. [134] employed two-dimensional DIC to measure surface displacements over a localized area of a concrete dam during turbine start-up and operation, capturing deformation patterns induced by dynamic loads and showing good agreement with independent geodetic measurements. In this context, DIC can exploit the natural texture of concrete surfaces, avoiding the need for artificial speckle patterns.

Despite the relatively small number of dam-specific studies currently available, DIC could potentially support two distinct monitoring objectives in dam engineering. First, it may facilitate damage-oriented inspections of crack-prone regions previously identified through numerical stress analyses, supporting localized assessment of strain concentrations and crack development. Second, it may serve applications focused on displacement monitoring in localized structural regions, such as areas surrounding construction or contraction joints, where measured surface displacement fields under operational loading could support substructuring strategies in FE modelling of dams [135].

However, the practical application of DIC for long-term dam monitoring remains subject to several operational constraints. In particular, repeatable measurements require adequate control of the camera positioning, although different UAV-based strategies may partially address this issue. A first possible approach is the deployment of stereo-camera systems on UAV platforms, where moderate repositioning variability can be tolerated provided that the stereo configuration remains calibrated and the target is acquired within a suitable viewing geometry. In localized repeated-inspection scenarios, permanent fiducial markers may also be required to facilitate repeated identification of the monitored area and ensure consistent registration between inspections, as demonstrated in UAV-based DIC applications for bridge monitoring [136]. A second alternative consists in using a single UAV-mounted camera, with the camera pose reconstructed through structure-from-motion algorithms exploiting stable external reference features visible in the surroundings, although this approach is only applicable where such features are available and sufficiently persistent over time [137].

Nevertheless, outdoor DIC measurements remain sensitive to environmental factors, particularly non-uniform illumination, moving shadows, and varying solar exposure, which may introduce localized disturbances and reduce measurement accuracy. Recent studies have shown that brightness-correction techniques, including Retinex- and frequency-based compensation approaches, can partially mitigate these effects by reducing measurement uncertainty under variable shadow conditions [138]. In periodic monitoring campaigns, additional practical precautions, such as scheduling acquisitions under comparable weather and lighting conditions, may improve repeatability. Moreover, although concrete surfaces may sometimes provide sufficient natural texture for correlation, localized high-accuracy measurements may still require artificial speckle preparation or surface treatment, particularly in smooth regions.

Consequently, DIC currently appears more suited to localized inspections and periodic targeted monitoring surveys than to fully autonomous continuous long-term deployment over large dam surfaces.

Recent research has also explored the integration of DIC with AI techniques. As reviewed by Sadeghian et al. [139], full-field displacement and strain maps obtained from DIC can be processed using CNN-based algorithms for automated damage identification, analogously to the analysis of RGB images. In particular, several studies have applied semantic segmentation architectures to DIC-derived strain maps, using FE-generated synthetic datasets for training. This strategy exploits simulated full-field responses associated with predefined damage scenarios, thereby allowing supervised learning without requiring extensive manual annotation of experimental strain maps [140,141]. A representative U-Net-based architecture employed for this purpose is illustrated in Fig. 12.

Another perspective involves the direct integration of DL into the DIC process itself. In this case, CNNs are employed to estimate displacement and strain fields from the acquired images, effectively replacing traditional correlation-based algorithms [142].

Overall, although applications of DIC and AI-enhanced DIC to dam monitoring remain limited, current studies indicate significant potential for localized deformation and damage assessment, provided that practical deployment challenges are addressed through further experimental investigation and field-scale validation.

5.3. Full-field-informed numerical modelling

The full-field techniques reviewed in the previous sections provide spatially distributed measurements of physically meaningful quantities, such as displacements, strains, and temperature fields, that can be directly exploited to calibrate and update physics-based numerical models.

In the context of dams, literature examples indicate that full-field displacement measurements are most commonly obtained through remote sensing techniques, such as ground-based or UAV-based radar and LiDAR systems. By reconstructing spatially distributed displacement fields from distance variations via temporal differencing and geometric processing, these measurements provide a richer description of the structural response than point-wise instrumentation and lead to better-conditioned inverse problems. This benefit has been demonstrated in early FE-based diagnostic studies combining radar-derived displacement data with gradient-based optimization techniques [143,144].

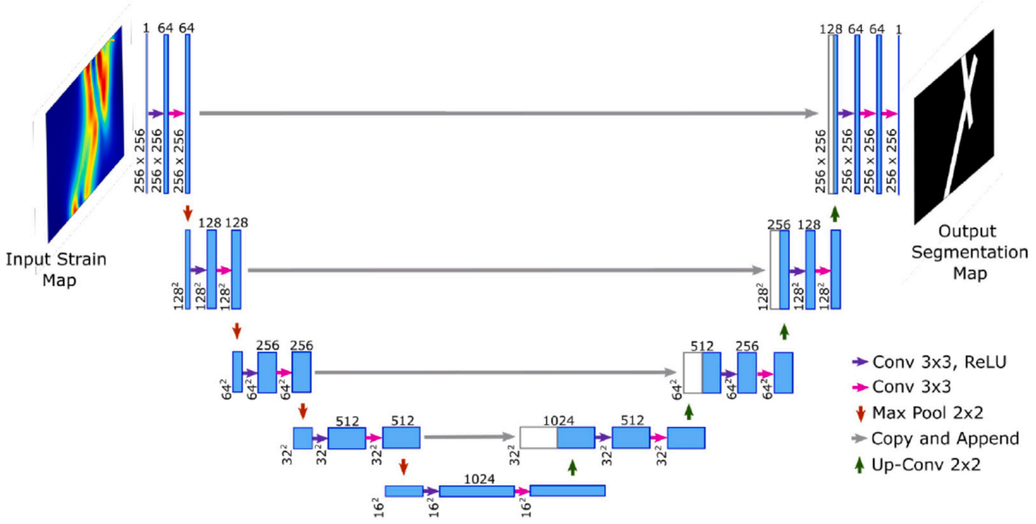

Fig. 12. U-Net-based architecture for semantic segmentation of DIC-derived strain maps used for damage identification.
Source: Reproduced from [141].

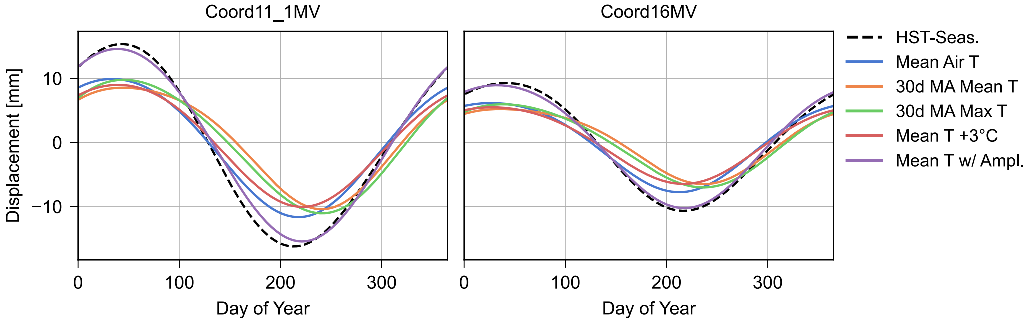

Fig. 13. Comparison between seasonal displacement components at two monitoring points of an Italian dam over one year: numerical simulations under different assumed thermal conditions (coloured lines) versus the seasonal component derived from HST decomposition of monitoring data (black dashed line).
Source: Adapted from [25].

For concrete dams, displacement interpretation is strongly affected by thermal effects, which can induce deformations comparable to hydrostatic loading. Inverse analyses of mechanical parameters therefore typically follow two strategies. A first approach relies on preliminary component separation, most commonly through HST-based statistical decompositions retaining only the hydrostatic component for calibration [145–147]. Such decomposition also enables the separate calibration of parameters associated with the thermal response. Fig. 13 illustrates this aspect by comparing seasonal displacement curves at two monitoring points of an Italian dam, obtained from numerical simulations under different assumed thermal conditions, with the seasonal component extracted from monitoring data using the HST method.

A more physically consistent alternative explicitly incorporates thermal inputs within the FE model, provided that suitable thermal information is available [148,149]. In this context, full-field temperature measurements from IRT (especially when acquired repeatedly over time) offer spatially resolved thermal data rarely accessible with pointwise sensors and enable more realistic thermal boundary conditions in inverse analyses.

From a spatial perspective, measurement resolution governs the scale of inverse analyses: radar and LiDAR capture global displacement patterns, whereas DIC provides high-resolution displacement and strain data over limited regions of interest, supporting inverse analyses focused on mechanically critical zones such as construction joints or dam–foundation interfaces.

From a computational perspective, the increasing complexity of inverse analyses has motivated the use of ML techniques as surrogate models. Rather than replacing the physical simulations, these approaches aim at approximating the forward structural response to drastically reduce the computational cost associated with repeated model evaluations during optimization.

Representative studies have explored surrogate-based inverse analyses using different regression models to approximate the numerical response, including NN [150], polynomial response surfaces combined with genetic algorithms [151], RBF models coupled with swarm-based optimization [152], and kernel-based regressors optimized through metaheuristic strategies [145]. As an alternative to surrogate-assisted optimization, direct ML inversion based on PINNs has recently been proposed. This approach performs parameter identification from sparse displacement data through explicit physical constraints, although current applications remain largely limited to two-dimensional formulation [153].

A comparative overview of the reviewed inverse-analysis approaches is provided in Table 5. While these computational strategies have so far been mainly applied to global-scale inverse problems using pointwise monitoring data, their underlying principles are not inherently scale-dependent. In this perspective, the combination of DIC-based displacement fields with surrogate-assisted inverse analysis represents a promising strategy for identifying local mechanical properties at joints, interfaces, or other mechanically critical regions [154,155].

Table 5
Overview of inverse and diagnostic approaches for dam monitoring based on displacement data.

Reference	Data used for inversion	Thermal effects	Type of inversion	Objective
Ardito et al. [143]	Full-field pseudo-experimental displacements (radar)	Not considered	Direct FEM + Gradient based opt.	Diagnosis
Ardito and Cocchetti [144]	Full-field pseudo-experimental displacements (radar)	Not considered	Direct FEM + Opt.	Diagnosis
Ardito et al. [148]	Pseudo-experimental displacements (radar)	Modelled in FEM	Direct thermo-mechanical FEM + Opt.	Diagnosis
Yu et al. [150]	FEM-generated displacements (pseudo-experimental)	Not considered	Surrogate-based (NN) + Opt.	Calibration
Yao et al. [151]	Point-wise monitoring displacements	Modelled in FEM	Surrogate-based (polynomial RSM) + Opt.	Calibration
Dou et al. [152]	FEM-generated displacements (pseudo-experimental)	Not considered	Surrogate-based (RBF) + Opt.	Diagnosis
Kang et al. [145]	Point-wise monitoring displacements	Separated via HST	Surrogate-based (KELM)+ Opt.	Calibration
Chen et al. [146]	Point-wise monitoring displacements	Separated via HST	Direct FEM + Opt.	Calibration
Yang et al. [147]	Point-wise monitoring displacements	Separated via HST	Direct FEM + Opt.	Calibration
Nguyen-Tuan et al. [149]	Displacements, temperatures, and uplift pressures	Modelled in FEM	Direct FEM + Opt.	Diagnosis

6. Conclusions

This review examined how AI is currently used in dam safety monitoring by distinguishing three main application areas: continuous point-wise monitoring, vision-based inspection using RGB images, and the interpretation of optical full-field data, with a particular focus on IRT and DIC, including their use within surrogate-assisted FE model updating. For each area, the analysis considered the nature of the available data, the specific ML methodologies adopted, and the type of information that can realistically be extracted.

For continuous point-wise monitoring, data-driven models have demonstrated strong performance as predictive tools and anomaly detectors. However, anomaly detection rarely translates directly into diagnosis, as most approaches provide limited insight into the underlying structural mechanisms unless they are explicitly trained on data derived from damage simulations. Computer vision methods perform a complementary task by enabling the direct identification of damage in images. They have significantly accelerated inspection procedures and improved their accuracy, particularly for the detection of cracks and surface deterioration. However, these approaches remain inherently focused on surface manifestations.

Optical full-field techniques, such as IRT and DIC, occupy an intermediate position between image-based inspection and numerical modelling. By providing spatially distributed fields of physically meaningful quantities, these methodologies can improve the conditioning of inverse problems and support the calibration of local mechanical parameters that influence the overall structural behaviour. Despite this potential, the combined use of full-field measurements with data-driven interpretation and FE model-based updating remains largely underexplored.

Overall, the reviewed literature indicates that current research efforts are primarily focused on improving performance within individual tasks, depending on the specific input data available, while interactions between different monitoring sources remain limited. Some initial forms of interaction can be observed, such as the combined use of RGB and infrared images for surface damage detection, or the integration of spatially distributed environmental information with point-wise measurements.

Recent advances in the use of LLMs within multimodal AI suggest a potential evolution toward a more integrated and generalizable interpretation of dam monitoring data. In particular, these approaches may help overcome three main limitations identified in this review.

First, they offer improved generalization capabilities beyond the training domain, including zero-shot and few-shot settings, which are particularly relevant for point-wise monitoring, where extrapolation and rare events remain critical challenges, and for vision-based inspection, where labelled examples of damage are inherently limited.

Second, their ability to process heterogeneous data provides a pathway toward a unified interpretation of multiple monitoring sources, including time series measurements, RGB images, and full-field maps. In this context, LLM-based systems can serve as orchestrators of heterogeneous analytical components, using agent-based or tool-calling approaches to integrate specialized methods, including regression and computer vision techniques. In addition, their capability to incorporate unstructured information, such as inspection reports, historical records, and technical documentation, enables the integration of contextual knowledge that is typically not captured by conventional monitoring workflows. This supports the combination of different data sources and modelling strategies, moving from fragmented analyses toward a more comprehensive understanding.

Third, these approaches may also facilitate the interaction between data-driven models and FE or physics-informed modelling, supporting a more consistent link between observations and structural behaviour. However, physics-informed methods such as PINNs, although conceptually well suited for this purpose, still require further investigation and validation before being applied to complex dam-scale problems.

The advancement of these emerging approaches remains closely dependent on the availability of diverse and representative monitoring data, which continues to be one of the main limitations in dam engineering. Improving data accessibility and establishing robust validation practices therefore represents a key step toward the deployment of applicable and trustworthy AI tools.

CRediT authorship contribution statement

Caterina Nogara: Conceptualization, Methodology, Investigation, Formal analysis, Data curation, Visualization, Writing – original draft, Writing – review & editing. **Gabriella Bolzon:** Conceptualization, Investigation, Validation, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

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.

Data availability

Data will be made available on request.

References

[1] N. Adamo, N. Al-Ansari, V. Sissakian, J. Laue, S. Knutsson, Dam safety: Technical problems of ageing concrete dams, J. Earth Sci. Geotech. Eng. 10 (6) (2020) 241–279.

[2] D. Perera, V. Smakhtin, S. Williams, T. North, R. Curry, Ageing water storage infrastructure: An emerging global risk, 2021, http://dx.doi.org/10.13140/RG.2.2.29149.44003.

[3] J. Fluixà-Sanmartín, L. Altarejos-García, A. Morales-Torres, I. Escuder-Bueno, Review article: Climate change impacts on dam safety, Nat. Hazards Earth Syst. Sci. 18 (9) (2018) 2471–2488, http://dx.doi.org/10.5194/nhess-18-2471-2018.

[4] J. Fluixà-Sanmartín, A. Morales-Torres, I. Escuder-Bueno, J. Paredes-Arquiola, Quantification of climate change impact on dam failure risk under hydrological scenarios: a case study from a spanish dam, Nat. Hazards Earth Syst. Sci. 19 (10) (2019) 2117–2139, http://dx.doi.org/10.5194/nhess-19-2117-2019.

[5] A. Moreno-Rodenas, J.D. Mantilla-Jones, D. Valero, Age, climate and economic disparities drive the current state of global dam safety, Nat. Water 3 (3) (2025) 284–295, http://dx.doi.org/10.1038/s44221-025-00402-1.

[6] ICOLD, Automated dam monitoring systems: Guidelines and case histories, Technical Report B-118, ICOLD, 2000.

[7] ICOLD, Dam surveillance guide, Technical Report B-158, ICOLD, 2012.

[8] C.R. Farrar, K. Worden, An introduction to structural health monitoring, Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci. 365 (1851) (2007) 303–315, http://dx.doi.org/10.1098/rsta.2006.1928.

[9] C. Boller, F.-K. Chang, Y. Fujino, Encyclopedia of Structural Health Monitoring, John Wiley & Sons, 2009, http://dx.doi.org/10.1002/9780470061626.

[10] F. Salazar, M.Á. Toledo, J.M. González, E. Oñate, Early detection of anomalies in dam performance: A methodology based on boosted regression trees, Struct. Control. Health Monit. 24 (11) (2017) e2012, http://dx.doi.org/10.1002/stc.2012.

[11] M. Azimi, A.D. Eslamlou, G. Pekcan, Data-driven structural health monitoring and damage detection through deep learning: State-of-the-art review, Sensors 20 (10) (2020) 2778, http://dx.doi.org/10.3390/s20102778.

[12] M.A. Hariri-Ardebili, G. Mahdavi, L.K. Nuss, U. Lall, The role of artificial intelligence and digital technologies in dam engineering: Narrative review and outlook, Eng. Appl. Artif. Intell. 126 (2023) 106813, http://dx.doi.org/10.1016/j.engappai.2023.106813.

[13] W. Cao, X. Wu, J. Li, F. Kang, A review of artificial intelligence in dam engineering, J. Infrastruct. Intell. Resil. 4 (1) (2025) 100122, http://dx.doi.org/10.1016/j.iintel.2024.100122.

[14] P. Bukenya, P. Moyo, H. Beushausen, C. Oosthuizen, Health monitoring of concrete dams: A literature review, J. Civ. Struct. Health Monit. 4 (4) (2014) 235–244, http://dx.doi.org/10.1007/s13349-014-0079-2.

[15] F. Salazar, R. Morán, M.Á. Toledo, E. Oñate, Data-based models for the prediction of dam behaviour: A review and some methodological considerations, Arch. Comput. Methods Eng. 24 (1) (2017) 1–21, http://dx.doi.org/10.1007/s11831-015-9157-9.

[16] X. Liu, Z. Li, L. Sun, E.Y. Khailah, J. Wang, W. Lu, A critical review of statistical model of dam monitoring data, J. Build. Eng. 80 (2023) 108106, http://dx.doi.org/10.1016/j.jobe.2023.108106.

[17] G. Prakash, R. Dugalam, M. Barbosh, A. Sadhu, Recent advancement of concrete dam health monitoring technology: A systematic literature review, Structures 44 (2022) 766–784, http://dx.doi.org/10.1016/j.istruc.2022.08.021.

[18] Z. Li, E.Y. Khailah, X. Liu, J. Liang, Exploring purpose-driven methods and a multifaceted approach in dam health monitoring data utilization, Buildings 15 (15) (2025) 2803, http://dx.doi.org/10.3390/buildings15152803.

[19] H. Najih, A. Aboulhassane, O.E.K. Moustachi, Contributions of artificial intelligence to dam monitoring: A literature review, Asian J. Civ. Eng. 26 (12) (2025) 4925–4940, http://dx.doi.org/10.1007/s42107-025-01474-w.

[20] Z. Peng, L. Li, D. Liu, S. Zhou, Z. Liu, A comprehensive survey on visual perception methods for intelligent inspection of high dam hubs, Sensors 24 (16) (2024) 5246, http://dx.doi.org/10.3390/s24165246.

[21] Y. Hamishebahar, H. Guan, S. So, J. Jo, A comprehensive review of deep learning-based crack detection approaches, Appl. Sci. 12 (3) (2022) 1374, http://dx.doi.org/10.3390/app12031374.

[22] F. Li, D. Luo, D. Niu, Data-intelligence driven methods for durability, damage diagnosis and performance prediction of concrete structures, Commun. Eng. 4 (1) (2025) 100, http://dx.doi.org/10.1038/s44172-025-00431-4.

[23] K. de Rubertis, Instrumentation and measurement tools, in: Monitoring Dam Performance, American Society of Civil Engineers, 2018, pp. 57–208, http://dx.doi.org/10.1061/9780784414828.ch05.

[24] K. de Rubertis, Planning and implementing a monitoring program, in: Monitoring Dam Performance, American Society of Civil Engineers, 2018, pp. 35–55, http://dx.doi.org/10.1061/9780784414828.ch04.

[25] C. Nogara, Machine Learning Tools for Structural Health Monitoring of Dams (Ph.D. thesis), Politecnico di Milano, Department of Civil and Environmental Engineering, Milan, Italy, 2025, URL https://hdl.handle.net/10589/244278.

[26] S. Ferry, G. Willm, Méthodes d'analyse et de surveillance des déplacements observés par le moyen de pendules dans les barrages, in: Proceedings of the 6th International Congress on Large Dams, New York, USA, 1958, pp. 1179–1201.

[27] P. Léger, M. Leclerc, Hydrostatic, temperature, time-displacement model for concrete dams, J. Eng. Mech. (ASCE) 133 (2007) 267–277, http://dx.doi.org/10.1061/(ASCE)0733-9399(2007)133:3(267).

[28] J. Mata, A.T.D. Castro, J.S.D. Costa, Constructing statistical models for arch dam deformation, Struct. Control. Health Monit. 21 (2014) 423–437, http://dx.doi.org/10.1002/stc.1575.

[29] M. Tatin, M. Briffaut, F. Dufour, A. Simon, J.P. Fabre, Thermal displacements of concrete dams: Accounting for water temperature in statistical models, Eng. Struct. 91 (2015) 26–39, http://dx.doi.org/10.1016/J.ENGSTRUCT.2015.01.047.

[30] M. Tatin, M. Briffaut, F. Dufour, A. Simon, J.P. Fabre, Statistical modelling of thermal displacements for concrete dams: Influence of water temperature profile and dam thickness profile, Eng. Struct. 165 (2018) 63–75, http://dx.doi.org/10.1016/j.engstruct.2018.03.010.

[31] I. Penot, B. Daumas, J.-P. Fabre, Monitoring behaviour, Int. Water Power Dam Constr. 57 (2005) 24–27.

[32] A. Simon, Behaviour prediction of a concrete arch dam: Data-based models used by the formulator of the theme a in an industrial context, in: M. Klun, A.K. zanowski, N. Humar (Eds.), Proceedings of the 16th International Benchmark Workshop on Numerical Analysis of Dams, SLOCOLD, Ljubljana, Slovenia, 2024.

[33] M.E.M. Billah, C. Ulrich, T. Andrian, A coupled statistical and numerical approach for the arch dam monitoring, in: M. Klun, A.K. zanowski, N. Humar (Eds.), Proceedings of the 16th International Benchmark Workshop on Numerical Analysis of Dams, SLOCOLD, Ljubljana, Slovenia, 2024.

[34] M. Gomez, J. Cunha, A. Paixão, R. Fernandes, Behaviour prediction of a concrete arch dam implemented with an HTT-fem hybrid model, in: M. Klun, A.K. zanowski, N. Humar (Eds.), Proceedings of the 16th International Benchmark Workshop on Numerical Analysis of Dams, SLOCOLD, Ljubljana, Slovenia, 2024.

[35] M. Azevedo, S. Leitão, B. Farinha, Castilho, Behaviour prediction of a concrete arch dam: Finite element modelling and models of separation of effects, in: M. Klun, A.K. zanowski, N. Humar (Eds.), Proceedings of the 16th International Benchmark Workshop on Numerical Analysis of Dams, SLOCOLD, Ljubljana, Slovenia, 2024.

[36] J. Mata, Interpretation of concrete dam behaviour with artificial neural network and multiple linear regression models, Eng. Struct. 33 (2011) 903–910.

[37] G. Xu, Application of RBF neural network in dam deformation prediction, Appl. Mech. Mater. 675 (2014) 261–264.

[38] J. Cheng, Y. Xiong, Application of extreme learning machine combination model for dam displacement prediction, Procedia Comput. Sci. 107 (2017) 373–378.

[39] Q. Ren, H. Li, M. Li, J. Zhang, T. Kong, Towards online monitoring of concrete dam displacement subject to time-varying environments: An improved sequential learning approach, Adv. Eng. Informatics 55 (2023) 101881.

[40] Y. Zhang, W. Zhang, Y. Li, L. Wen, X. Sun, AF-os-ELM-mve: A new online sequential extreme learning machine of dam safety monitoring model for structure deformation estimation, Adv. Eng. Informatics 60 (2024) 102345, http://dx.doi.org/10.1016/j.aei.2023.102345.

[41] V. Ranković, N. Grujović, D. Divac, N. Milivojević, Development of support vector regression identification model for prediction of dam structural behaviour, Struct. Saf. 48 (2014) 33–39, http://dx.doi.org/10.1016/j.strusafe.2014.02.004.

[42] H. Su, Z. Chen, Z. Wen, Performance improvement method of support vector machine-based model monitoring dam safety, Struct. Control. Health Monit. 23 (2016) 252–266, http://dx.doi.org/10.1002/stc.1767.

[43] S. Chen, C. Gu, C. Lin, Y. Wang, M.A. Hariri-Ardebili, Prediction, monitoring, and interpretation of dam leakage flow via adaptative kernel extreme learning machine, Measurement 166 (2020) 108161, http://dx.doi.org/10.1016/j.measurement.2020.108161.

[44] Y. Lin, Q. Chen, M. Hariri-Ardebili, Interpretable KELM data-driven model for the prediction and monitoring of arch dam behaviour, in: M. Klun, A.K. zanowski, N. Humar (Eds.), Proceedings of the 16th International Benchmark Workshop on Numerical Analysis of Dams, SLOCOLD, Ljubljana, Slovenia, 2024.

[45] K. Yao, Z. Wen, L. Yang, J. Chen, H. Hou, H. Su, A multipoint prediction model for nonlinear displacement of concrete dam, Computer-Aided Civ. Infrastruct. Eng. 37 (14) (2022) 1932–1952, http://dx.doi.org/10.1111/mice.12911.

[46] F. Salazar, M.A. Toledo, E. Oñate, R. Morán, An empirical comparison of machine learning techniques for dam behaviour modelling, Struct. Saf. 56 (2015) 9–17, http://dx.doi.org/10.1016/j.strusafe.2015.05.001.

[47] A.M. Babadi, H. Mirzabozorg, K. Baharan, Predictive modeling of concrete arch dam behavior: Evaluating the efficacy of random forest and radial basis function networks, AI Civ. Eng. 4 (1) (2025) 22, http://dx.doi.org/10.1007/s43503-025-00071-9.

[48] F. Salazar, M.T. Toledo, E. Oñate, B. Suárez, Interpretation of dam deformation and leakage with boosted regression trees, Eng. Struct. 119 (2016) 230–251, http://dx.doi.org/10.1016/j.engstruct.2016.04.012.

[49] F. Salazar, M. Irazábal, J. Vincente, Prediction and interpretation of dam response with boosted regression trees, in: M. Klun, A.K. zanowski, N. Humar (Eds.), Proceedings of the 16th International Benchmark Workshop on Numerical Analysis of Dams, SLOCOLD, Ljubljana, Slovenia, 2024.

[50] C. Lin, T. Li, S. Chen, X. Liu, C. Lin, S. Liang, Gaussian process regression-based forecasting model of dam deformation, Neural Comput. Appl. 31 (2019) http://dx.doi.org/10.1007/s00521-019-04375-7.

[51] S. Chen, C. Gu, C. Lin, K. Zhang, Y. Zhu, Multi-kernel optimized relevance vector machine for probabilistic prediction of concrete dam displacement, Eng. Comput. 37 (2021) http://dx.doi.org/10.1007/s00366-019-00924-9.

[52] W. Xi, J. Yang, J. Song, X. Qu, Deep learning model of concrete dam deformation prediction based on CNN, IOP Conf. Ser.: Earth Environ. Sci. 580 (2020) 012042, http://dx.doi.org/10.1088/1755-1315/580/1/012042.

[53] J. Pan, W. Liu, C. Liu, J. Wang, Convolutional neural network-based spatiotemporal prediction for deformation behavior of arch dams, Expert Syst. Appl. 232 (2023) 120835, http://dx.doi.org/10.1016/j.eswa.2023.120835.

[54] J. Mata, F. Salazar, J. Barateiro, A. Antunes, Validation of machine learning models for structural dam behaviour interpretation and prediction, Water 13 (19) (2021) http://dx.doi.org/10.3390/w13192717.

[55] P. Alocén, M.Á. Fernández-Centeno, M.Á. Toledo, Greedy weighted stacking of machine learning models for optimizing dam deformation prediction, Water 16 (9) (2024) http://dx.doi.org/10.3390/w16091235.

[56] P. Alocén, M.Á. Fernández-Centeno, M.Á. Toledo, Prediction of concrete dam deformation through the combination of machine learning models, Water 14 (7) (2022) http://dx.doi.org/10.3390/w14071133.

[57] M. Fernández-Centeno, P. Alocén, R. Toledo, Prediction of dam behaviour based on machine learning methods, in: M. Klun, A.K. zanowski, N. Humar (Eds.), Proceedings of the 16th International Benchmark Workshop on Numerical Analysis of Dams, SLOCOLD, Ljubljana, Slovenia, 2024.

[58] E. Catalano, M. Stucchi, Behaviour prediction of a concrete arch dam, in: M. Klun, A.K. zanowski, N. Humar (Eds.), Proceedings of the 16th International Benchmark Workshop on Numerical Analysis of Dams, SLOCOLD, Ljubljana, Slovenia, 2024.

[59] J. Mata, F. Miranda, A. Antunes, X. Romão, J.P. Santos, Characterization of relative movements between blocks observed in a concrete dam and definition of thresholds for novelty identification based on machine learning models, Water 15 (2) (2023) http://dx.doi.org/10.3390/w15020297.

[60] N. Silva-Cancino, F. Salazar, J. Irazábal, J. Mata, Adaptive warning thresholds for dam safety: A KDE-based approach, Infrastructures 10 (7) (2025) http://dx.doi.org/10.3390/infrastructures10070158.

[61] Q. Ren, M. Li, R. Kong, Y. Shen, S. Du, A hybrid approach for interval prediction of concrete dam displacements under uncertain conditions, Eng. Comput. 39 (2) (2023) 1285–1303, http://dx.doi.org/10.1007/s00366-021-01515-3.

[62] Q. Ren, M. Li, Y. Shen, A new interval prediction method for displacement behavior of concrete dams based on gradient boosted quantile regression, Struct. Control. Health Monit. 29 (1) (2022) e2859, http://dx.doi.org/10.1002/stc.2859.

[63] J. Mata, N. Serra, Behaviour prediction of a concrete arch dam combining NN and MLR models – proposal for the 16th ICOLD BW, in: M. Klun, A.K. zanowski, N. Humar (Eds.), Proceedings of the 16th International Benchmark Workshop on Numerical Analysis of Dams, SLOCOLD, Ljubljana, Slovenia, 2024.

[64] F. Salazar, M.Á. Toledo, J.M. González, E. Oñate, Early detection of anomalies in dam performance: A methodology based on boosted regression trees, Struct. Control. Health Monit. 24 (11) (2017) e2012, http://dx.doi.org/10.1002/stc.2012.

[65] G.E.P. Box, G.M. Jenkins, Time series analysis: Forecasting and control, Revised ed., Holden-Day, San Francisco, 1976.

[66] Y.H. Chen, J.G. Zou, B. Li, T. Wang, Application of multiplicative seasonal ARIMA model in dam displacement monitoring, J Geomat 39.2 (2014) 35–38.

[67] A. Corigliano, F. Moscariello, L'Aurora, Pasqualato, Behaviour prediction of a concrete arch dam for the 2022 icold benchmark, in: M. Klun, A.K. zanowski, N. Humar (Eds.), Proceedings of the 16th International Benchmark Workshop on Numerical Analysis of Dams, SLOCOLD, Ljubljana, Slovenia, 2024.

[68] K.-T. Bui, X. Yangxuan, C. Doan, X.K. Do, T.V. Tran, D.S. Mai, Application of statistic model and backpropagation neural network to analyzing and forecasting hydropower dam displacement, VNU J. Sci.: Earth Environ. Sci. 37 (2021) http://dx.doi.org/10.25073/2588-1094/vnuees.4529.

[69] Y. Su, J. Fu, W. Lin, C. Lin, X. Lai, X. Xie, Dam deformation monitoring model based on deep learning and split conformal quantile prediction, Appl. Sci. 15 (4) (2025) 1960, http://dx.doi.org/10.3390/app15041960.

[70] C. Lin, X. Wang, Y. Su, T. Zhang, C. Lin, Deformation forecasting of pulp-masonry arch dams via a hybrid model based on CEEMDAN considering the lag of influencing factors, J. Struct. Eng. (ASCE) 148 (2022) http://dx.doi.org/10.1061/(ASCE)ST.1943-541X.0003356.

[71] M. Pirker, T. Zenz, Hybrid analysis of an arch dam with quantile regression neural network, in: M. Klun, A.K. zanowski, N. Humar (Eds.), Proceedings of the 16th International Benchmark Workshop on Numerical Analysis of Dams, SLOCOLD, Ljubljana, Slovenia, 2024.

[72] J.-A. Goulet, L.H. Nguyen, S. Amiri, Tractable approximate Gaussian inference for Bayesian neural networks, J. Mach. Learn. Res. 22 (2021) 1–23.

[73] S. Deka, T. Vuong, J.-M. Goulet, M. Côté, P. Miquel, Dam behaviour prediction using an ensemble of Bayesian dynamic linear model and Bayesian LSTM networks, in: M. Klun, A.K. zanowski, N. Humar (Eds.), Proceedings of the 16th International Benchmark Workshop on Numerical Analysis of Dams, SLOCOLD, Ljubljana, Slovenia, 2024.

[74] W. Liu, J. Pan, Y. Ren, Z. Wu, J. Wang, Coupling prediction model for long-term displacements of arch dams based on long short-term memory network, Struct. Control. Health Monit. 27 (7) (2020) e2548, http://dx.doi.org/10.1002/stc.2548.

[75] G. Hua, S. Wang, M. Xiao, S. Hu, Research on the uplift pressure prediction of concrete dams based on the CNN–GRU model, Water 15 (2) (2023) 319, http://dx.doi.org/10.3390/w15020319.

[76] Y. Wei, Q. Li, Y. Hu, Y. Wang, X. Zhu, Y. Tan, C. Liu, L. Pei, Deformation prediction model based on an improved CNN–LSTM model for the first impoundment of super-high arch dams, J. Civ. Struct. Health Monit. 13 (2) (2023) 431–442, http://dx.doi.org/10.1007/s13349-022-00640-x.

[77] W. Cao, Z. Wen, Y. Feng, S. Zhang, H. Su, A multi-point joint prediction model for high-arch dam deformation considering spatial and temporal correlation, Water 16 (10) (2024) 1388, http://dx.doi.org/10.3390/w16101388.

[78] B. Xu, Z. Chen, H. Su, H. Zhang, A deep learning method for predicting the displacement of concrete arch dams considering the effect of cracks, Adv. Eng. Inform. 62 (2024) 102574, http://dx.doi.org/10.1016/j.aei.2024.102574.

[79] M. Li, M. Li, Q. Ren, H. Li, L. Song, DRLSTM: A dual-stage deep learning approach driven by raw monitoring data for dam displacement prediction, Adv. Eng. Inform. 51 (2022) 101510, http://dx.doi.org/10.1016/j.aei.2021.101510.

[80] B. Huang, F. Kang, J. Li, F. Wang, Displacement prediction model for high arch dams using long short-term memory based encoder-decoder with dual-stage attention considering measured dam temperature, Eng. Struct. 280 (2023) 115686, http://dx.doi.org/10.1016/j.engstruct.2023.115686.

[81] M. Li, Q. Ren, M. Li, Y. Chen, X. Ji, H. Liu, Efficient prediction uncertainty quantification in dam behavior monitoring with attention-based sequence-to-sequence learning, Appl. Soft Comput. 167 (2024) 112321, http://dx.doi.org/10.1016/j.asoc.2024.112321.

[82] K. Cho, B. van Merriënboer, D. Bahdanau, Y. Bengio, On the properties of neural machine translation: Encoder–decoder approaches, in: D. Wu, M. Carpuat, X. Carreras, E.M. Vecchi (Eds.), Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation, Association for Computational Linguistics, Doha, Qatar, 2014, pp. 103–111, http://dx.doi.org/10.3115/v1/W14-4012.

[83] F. Salazar, A. Conde, D.J. Vicente, Identification of dam behavior by means of machine learning classification models, in: G. Bolzon, D. Sterpi, G. Mazzà, A. Frigerio (Eds.), Numerical Analysis of Dams, Springer International Publishing, Cham, 2021, pp. 851–862, http://dx.doi.org/10.1007/978-3-030-51085-5-48.

[84] F. Salazar, A. Conde, J. Irazábal, D.J. Vicente, Anomaly detection in dam behaviour with machine learning classification models, Water 13 (17) (2021) http://dx.doi.org/10.3390/w13172387.

[85] B. Chen, T. Hu, Z. Huang, C. Fang, A spatio-temporal clustering and diagnosis method for concrete arch dams using deformation monitoring data, Struct. Health Monit. 18 (5–6) (2019) 1355–1371, http://dx.doi.org/10.1177/1475921718797949.

[86] J. Wang, H. Gu, B. Chen, C. Gu, Q. Zhang, Z. Xing, A spatio-temporal dam deformation zoning method considering non-uniform distribution of monitoring information, IEEE Access 9 (2021) 117615–117628, http://dx.doi.org/10.1109/ACCESS.2021.3106817.

[87] W. Lei, J. Wang, T. Ji, P. Li, Dam deformation early warning model based on cluster analysis and spatiotemporal data fusion, Measurement 204 (2022) 112109, http://dx.doi.org/10.1016/j.measurement.2022.112109.

[88] C. Nogara, F. Salazar, Detection and spatial localization of dam anomalies using autoencoder residuals, NDT E Int. 161 (2026) 103696, http://dx.doi.org/10.1016/j.ndteint.2026.103696.

[89] X. Shu, T. Bao, Y. Zhou, R. Xu, Y. Li, K. Zhang, Unsupervised dam anomaly detection with spatial–temporal variational autoencoder, Struct. Health Monit. 22 (1) (2023) 39–55, http://dx.doi.org/10.1177/14759217211073301.

[90] Y. Wang, G. Liu, Self-supervised dam deformation anomaly detection based on temporal–spatial contrast learning, Sensors 24 (2024) 5858, http://dx.doi.org/10.3390/s24175858.

[91] C. Tutivén, L. Moyón, F. Salazar, A robust approach for anomaly detection using 1D convolutional siamese neural networks to enhance structural health monitoring in dams, Struct. Health Monit. (2025) http://dx.doi.org/10.1177/14759217251372614.

[92] X. Kang, Y. Li, Y. Zhang, N. Ma, L. Wen, Anomaly detection in concrete dam using memory-augmented autoencoder and generative adversarial network, Autom. Constr. 168 (2024) 105794, http://dx.doi.org/10.1016/j.autcon.2024.105794.

[93] S. Li, B. Zhang, J. Zheng, Z. Liu, X. Zhang, G. Ye, C. Yang, L. Wang, H. Tang, Time-LLM-based multisource and multihorizon deformation forecasting for dam safety monitoring, Results Eng. 30 (2026) 110407, http://dx.doi.org/10.1016/j.rineng.2026.110407.

[94] C. Liu, R. Qu, C. Liu, J. Wang, J. Pan, LLM-based multi-agent system for dam structural health monitoring, Autom. Constr. 186 (2026) 106895, http://dx.doi.org/10.1016/j.autcon.2026.106895.

[95] F. Salazar, A. Simon, R. Malm, R. Hellgren, M. Klun, Behaviour prediction of a concrete arch dam – description and synthesis of theme a, in: Proceedings of the 16th International Benchmark Workshop on Numerical Analysis of Dams, Slovenian National Committee on Large Dams (SLOCOLD), Ljubljana, Slovenia, 2024, pp. 8–25.

[96] F. Salazar, J. Irazábal, N. Silva-Cancino, D.J. Vicente, Performance evaluation of hybrid approaches for predicting arch dam deformations, in: IOP Conference Series: Earth and Environmental Science, vol. 1575, (1) IOP Publishing, 2025, 012025, http://dx.doi.org/10.1088/1755-1315/1575/1/012025.

[97] M. Hajjar, E. Zappa, G. Bolzon, The state of the art and potentialities of UAV-based 3D measurement solutions in the monitoring and fault diagnosis of quasi-brittle structures, Sensors 25 (16) (2025) 5134, http://dx.doi.org/10.3390/s25165134.

[98] C. Feng, H. Zhang, H. Wang, S.W.S. Wang, Y.L.Y. Li, Automatic pixel-level crack detection on dam surface using deep convolutional network, Sensors 20 (7) (2020) 2069, http://dx.doi.org/10.3390/s20072069.

[99] R. Archana, P.S.E. Jeevaraj, Deep learning models for digital image processing: a review, Artif. Intell. Rev. 57 (1) (2024) 11, http://dx.doi.org/10.1007/s10462-023-10631-z.

[100] M. Trigka, E. Dritsas, A comprehensive survey of deep learning approaches in image processing, Sensors 25 (2) (2025) 531, http://dx.doi.org/10.3390/s25020531.

[101] A. Dede, H. Nunoo-Mensah, E. Tutu Tchao, A.S. Agbemenu, P.E. Adjei, F.A. Acheampong, J.J. Kponyo, Deep learning for efficient high-resolution image processing: A systematic review, Intell. Syst. Appl. 26 (2025) 200505, http://dx.doi.org/10.1016/j.iswa.2025.200505.

[102] L. Parente, E. Falvo, C. Castagnetti, F. Grassi, F. Mancini, P. Rossi, A. Capra, Image-based monitoring of cracks: Effectiveness analysis of an open-source machine learning-assisted procedure, J. Imaging 8 (2) (2022) 22, http://dx.doi.org/10.3390/jimaging8020022.

[103] H.C. Reis, V. Turk, M. Ustuner, C.M. Kaya Yildiz, R. Tatli, Post-seismic structural assessment: advanced crack detection through complex feature extraction using pre-trained deep learning and machine learning integration, Earth Sci. Informatics 18 (1) (2025) 133, http://dx.doi.org/10.1007/s12145-024-01574-2.

[104] B. Chen, H. Zhang, Y.L.Y. Li, S.W.S. Wang, H. Zhou, H. Lin, Quantify pixel-level detection of dam surface crack using deep learning, Meas. Sci. Technol. 33 (6) (2022) 065402, http://dx.doi.org/10.1088/1361-6501/ac4b8d.

[105] Z. Wu, Y. Tang, B. Hong, B. Liang, Y. Liu, Enhanced precision in dam crack width measurement: Leveraging advanced lightweight network identification for pixel-level accuracy, Int. J. Intell. Syst. 2023 (1) (2023) 9940881, http://dx.doi.org/10.1155/2023/9940881.

[106] S.Z.S. Zhao, F. Kang, J.L.J. Li, Intelligent segmentation method for blurred cracks and 3D mapping of width nephograms in concrete dams using UAV photogrammetry, Autom. Constr. 157 (2024) 105145, http://dx.doi.org/10.1016/j.autcon.2023.105145.

[107] J. Xu, C. Yuan, J. Gu, J. Liu, J. An, Q. Kong, Innovative synthetic data augmentation for dam crack detection, segmentation, and quantification, Struct. Health Monit. 22 (4) (2023) 2402–2426, http://dx.doi.org/10.1177/14759217221122318.

[108] U.S. Khan, M. Ishfaque, S.U.R. Khan, F. Xu, L. Chen, Y. Lei, Comparative analysis of twelve transfer learning models for the prediction and crack detection in concrete dams, Front. Struct. Civ. Eng. 18 (10) (2024) 1507–1523, http://dx.doi.org/10.1007/s11709-024-1090-2.

[109] J. Zhang, T. Bao, An improved ResNet-based algorithm for crack detection of concrete dams using dynamic knowledge distillation, Water 15 (15) (2023) 2839, http://dx.doi.org/10.3390/w15152839.

[110] X. Li, L.L.L. Li, Z. Liu, Z. Peng, S. Liu, S. Zhou, X. Chai, K. Jiang, Dam crack detection studies by UAV based on YOLO algorithm, in: 2023 2nd International Conference on Robotics, Artificial Intelligence and Intelligent Control, RAIIC, 2023, pp. 104–108, http://dx.doi.org/10.1109/RAIIC59453.2023.10281120.

[111] B.H.B. Huang, S.Z.S. Zhao, F. Kang, Image-based automatic multiple-damage detection of concrete dams using region-based convolutional neural networks, J. Civ. Struct. Health Monit. 13 (2) (2023) 413–429, http://dx.doi.org/10.1007/s13349-022-00650-9.

[112] Y.L.Y. Li, T. Bao, A real-time multi-defect automatic identification framework for concrete dams via improved YOLOv5 and knowledge distillation, J. Civ. Struct. Health Monit. 13 (6) (2023) 1333–1349, http://dx.doi.org/10.1007/s13349-023-00684-7.

[113] S.Z.S. Zhao, F. Kang, J.L.J. Li, Concrete dam damage detection and localisation based on YOLOv5s-HSC and photogrammetric 3D reconstruction, Autom. Constr. 143 (2022) 104555, http://dx.doi.org/10.1016/j.autcon.2022.104555.

[114] K. Hong, H.W.H. Wang, B. Yuan, T.W.T. Wang, Multiple defects inspection of dam spillway surface using deep learning and 3D reconstruction techniques, Buildings 13 (2) (2023) 285, http://dx.doi.org/10.3390/buildings13020285.

[115] S. Zhu, X.L.X. Li, G. Wan, H.W.H. Wang, S. Shao, P.S.P. Shi, Underwater dam crack image classification algorithm based on improved VanillaNet, Symmetry 16 (7) (2024) 845, http://dx.doi.org/10.3390/sym16070845.

[116] P.S.P. Shi, X. Fan, J. Ni, Z. Khan, M.L.M. Li, A novel underwater dam crack detection and classification approach based on sonar images, PLoS One 12 (6) (2017) 1–17, http://dx.doi.org/10.1371/journal.pone.0179627.

[117] Y.L.Y. Li, T. Bao, X. Huang, H. Chen, B. Xu, X. Shu, Y. Zhou, Q. Cao, J. Tu, R.W.R. Wang, K. Zhang, Underwater crack pixel-wise identification and quantification for dams via lightweight semantic segmentation and transfer learning, Autom. Constr. 144 (2022) 104600, http://dx.doi.org/10.1016/j.autcon.2022.104600.

[118] B.H.B. Huang, F. Kang, X.L.X. Li, S. Zhu, Underwater dam crack image generation based on unsupervised image-to-image translation, Autom. Constr. 163 (2024) 105430, http://dx.doi.org/10.1016/j.autcon.2024.105430.

[119] F. Kang, B.H.B. Huang, G. Wan, Automated detection of underwater dam damage using remotely operated vehicles and deep learning technologies, Autom. Constr. 171 (2025) 105971, http://dx.doi.org/10.1016/j.autcon.2025.105971.

[120] P. Guo, X. Tan, Y. Liu, Diffusion-driven generation of synthetic complex concrete crack images for segmentation tasks, Struct. Durab. Health Monit. 20 (1) (2026) http://dx.doi.org/10.32604/sdhm.2025.071317.

[121] S. Duan, X. Tan, P. Guo, Y. Guo, Y. Bao, The transformative roles of generative artificial intelligence in vision techniques for structural health monitoring: A state-of-the-art review, Adv. Eng. Inform. 68 (2025) 103719, http://dx.doi.org/10.1016/j.aei.2025.103719.

[122] M. Henriques, P. Ramos, Thermal imaging of concrete dam surfaces to support the control of the evolution of pathologies, in: Second International Dam World Conference, Lisbon, Portugal, 2015.

[123] Y. He, B. Deng, H. Wang, L. Cheng, K. Zhou, S. Cai, F. Ciampa, Infrared machine vision and infrared thermography with deep learning: A review, Infrared Phys. Technol. 116 (2021) 103754, http://dx.doi.org/10.1016/j.infrared.2021.103754.

[124] S. Pozzer, E. Rezazadeh Azar, F. Dalla Rosa, Z.M. Chamberlain Pravia, Semantic segmentation of defects in infrared thermographic images of highly damaged concrete structures, J. Perform. Constr. Facil. 35 (1) (2021) 04020131, http://dx.doi.org/10.1061/(ASCE)CF.1943-5509.0001541.

[125] Z.F. Wang, Y.F. Yu, J. Wang, J.Q. Zhang, H.L. Zhu, P. Li, L. Xu, H.N. Jiang, Q.M. Sui, L. Jia, J.P. Chen, Convolutional neural-network-based automatic dam-surface seepage defect identification from thermograms collected from UAV-mounted thermal imaging camera, Constr. Build. Mater. 323 (2022) 126416, http://dx.doi.org/10.1016/j.conbuildmat.2022.126416.

[126] S.Z.S. Zhao, F. Kang, L. He, J.L.J. Li, Y. Si, Y. Xu, Intelligent structural health monitoring and noncontact measurement method of small reservoir dams using UAV photogrammetry and anomaly detection, Appl. Sci. 14 (20) (2024) 9156, http://dx.doi.org/10.3390/app14209156.

[127] B.-Y. Wang, Detection method for surface personnel and leakage on small-scale earth-rock dams based on YOLOv8s, China Rural. Water Hydropower (2) (2025) 213–218, http://dx.doi.org/10.12396/znsd.240723.

[128] F. Zhu, Y. Gu, S. Wang, W. Zhou, Super-resolution study of infrared images of embankment dam leakage based on deep learning, in: Proceedings of the 2024 International Conference on Mathematics and Machine Learning, ICMML '24, Association for Computing Machinery, New York, NY, USA, 2025, pp. 132–140, http://dx.doi.org/10.1145/3708360.3708382.

[129] N. Saeed, N. King, Z. Said, M.A. Omar, Automatic defects detection in CFRP thermograms, using convolutional neural networks and transfer learning, Infrared Phys. Technol. 102 (2019) 103048, http://dx.doi.org/10.1016/j.infrared.2019.103048.

[130] Q. Luo, B. Gao, W.L. Woo, Y. Yang, Temporal and spatial deep learning network for infrared thermal defect detection, NDT E Int. 108 (2019) 102164, http://dx.doi.org/10.1016/j.ndteint.2019.102164.

[131] J. Yang, J. Wang, F. Jin, J. Pan, A physics-informed convolutional neural network for spatiotemporal temperature analysis of concrete dams, Eng. Appl. Artif. Intell. 150 (2025) 110624, http://dx.doi.org/10.1016/j.engappai.2025.110624.

[132] D. Feng, M.Q. Feng, Computer vision for structural health monitoring of civil infrastructure: From dynamic response measurement to damage detection, Eng. Struct. 156 (2018) 105–117, http://dx.doi.org/10.1016/j.engstruct.2017.11.018.

[133] D. Lei, Z. Huang, P. Bai, F. Zhu, Experimental research on impact damage of xiaowan arch dam model by digital image correlation, Constr. Build. Mater. 147 (2017) 168–173, http://dx.doi.org/10.1016/j.conbuildmat.2017.04.143.

[134] J. Zaczek-Peplinska, M. Kowalska, K. Malowany, M. Malesa, Application of digital image correlation and geodetic displacement measuring methods to monitor water dam behavior under dynamic load, J. Civ. Eng. Archit. 9 (12) (2015) http://dx.doi.org/10.17265/1934-7359/2015.12.011.

[135] P. Sengupta, S. Chakraborty, A state-of-the-art review on model reduction and substructuring techniques in finite element model updating for structural health monitoring applications, Arch. Comput. Methods Eng. 32 (5) (2025) 3031–3062, http://dx.doi.org/10.1007/s11831-025-10231-w.

[136] D. Reagan, A. Sabato, C. Niezrecki, Feasibility of using digital image correlation for unmanned aerial vehicle structural health monitoring of bridges, Struct. Health Monit. 17 (5) (2018) 1056–1072, http://dx.doi.org/10.1177/1475921717735326.

[137] F. Barros, P.J. Sousa, P.J. Tavares, P.M.G.P. Moreira, Digital image correlation with a moving camera using structure from motion calibration, Procedia Struct. Integr. 17 (2019) 986–991, http://dx.doi.org/10.1016/j.prostr.2019.08.131, 3rd International Conference on Structural Integrity, ICSI 2019, 2–5 September 2019, Funchal, Madeira, Portugal.

[138] M. Hajjar, E. Zappa, G. Bolzon, DIC uncertainty due to shadow variability in outdoor testing: Effects and mitigation actions, J. Phys.: Conf. Ser. 3063 (1) (2025) 012002, http://dx.doi.org/10.1088/1742-6596/3063/1/012002.

[139] M. Sadeghian, A. Palevicius, J. Sablinskas, P. Griskevicius, From pixels to predictions: Integrating machine learning and digital image correlation for damage identification in engineering materials, Materials 19 (1) (2026) 77, http://dx.doi.org/10.3390/ma19010077.

[140] Y. Wang, Q. Luo, H. Xie, Q. Li, G. Sun, Digital image correlation-based damage detection for CFRP laminates using machine learning-based image semantic segmentation, Int. J. Mech. Sci. 230 (2022) 107529, http://dx.doi.org/10.1016/j.ijmecsci.2022.107529.

[141] A. Pal, W. Meng, S. Nagarajaiah, Deep learning-based subsurface damage localization using full-field surface strains, Sensors 23 (17) (2023) 7445, http://dx.doi.org/10.3390/s23177445.

[142] R. Yang, Y. Li, D. Zeng, P. Guo, Deep DIC: Deep learning-based digital image correlation for end-to-end displacement and strain measurement, J. Mater. Process. Technol. 302 (2022) 117474, http://dx.doi.org/10.1016/j.jmatprotec.2021.117474.

[143] R. Ardito, P. Bartalotta, L. Ceriani, G. Maier, Diagnostic inverse analysis of concrete dams with statical excitation, J. Mech. Behav. Mater. 15 (2004) 381–390, http://dx.doi.org/10.1515/JMBM.2004.15.6.381.

[144] R. Ardito, G. Cocchetti, Statistical approach to damage diagnosis of concrete dams by radar monitoring: Formulation and a pseudo-experimental test, Eng. Struct. 28 (14) (2006) 2036–2045, http://dx.doi.org/10.1016/j.engstruct.2006.04.001.

[145] F. Kang, X. Liu, J. Li, H. Li, Multi-parameter inverse analysis of concrete dams using kernel extreme learning machines-based response surface model, Eng. Struct. 256 (2022) 113999, http://dx.doi.org/10.1016/j.engstruct.2022.113999.

[146] Y. Chen, C. Gu, B. Wu, C. Shao, Z. Wu, B. Dai, Inversion modeling of dam-zoning elasticity modulus for heightened concrete dam using ICS-IPSO algorithm, Math. Probl. Eng. 2019 (1) (2019) 9328326, http://dx.doi.org/10.1155/2019/9328326.

[147] L. Yang, H. Su, Z. Wen, Improved PLS and PSO methods-based back analysis for elastic modulus of dam, Adv. Eng. Softw. 131 (2019) 205–216, http://dx.doi.org/10.1016/j.advengsoft.2019.02.005.

[148] R. Ardito, G. Maier, G. Massalongo, Diagnostic analysis of concrete dams based on seasonal hydrostatic loading, Eng. Struct. 30 (11) (2008) 3176–3185, http://dx.doi.org/10.1016/j.engstruct.2008.04.008.

[149] L. Nguyen-Tuan, C. Könke, T. Lahmer, Damage identification using inverse analysis for 3D coupled thermo-hydro-mechanical problems, Comput. Struct. 196 (2018) 146–156, http://dx.doi.org/10.1016/j.compstruc.2017.11.008.

[150] Y. Yu, B. Zhang, H. Yuan, An intelligent displacement back-analysis method for earth-rockfill dams, Comput. Geotech. 34 (6) (2007) 423–434, http://dx.doi.org/10.1016/j.compgeo.2007.03.002.

[151] F. Yao, S. Guan, H. Yang, Y. Chen, H. Qiu, G. Ma, Q. Liu, Long-term deformation analysis of shuibuya concrete face rockfill dam based on response surface method and improved genetic algorithm, Water Sci. Eng. 12 (3) (2019) 196–204, http://dx.doi.org/10.1016/j.wse.2019.09.004.

[152] S. Dou, J. Li, F. Kang, Health diagnosis of concrete dams using hybrid FWA with RBF-based surrogate model, Water Sci. Eng. 12 (3) (2019) 188–195, http://dx.doi.org/10.1016/j.wse.2019.09.002.

[153] D. Luo, H. Mo, Q. Li, X. Jin, Solution of stress and deformation fields and inversion of material parameters for gravity dams based on physics-informed neural networks, J. Comput. Sci. 89 (2025) 102613, http://dx.doi.org/10.1016/j.jocs.2025.102613.

[154] G. Bolzon, A. Frigerio, M. Hajjar, C. Nogara, E. Zappa, Structural health assessment of existing dams based on non-destructive testing, physics-based models and machine learning tools, NDT & E Int. 150 (2025) 103271, http://dx.doi.org/10.1016/j.ndteint.2024.103271.

[155] G. Bolzon, C. Nogara, Open issues in structural health monitoring and local failure detection in concrete dams, in: Proceedings of the 2nd Latin-American Workshop on Structural Health Monitoring, 31(2), Santiago, Chile, 2026, http://dx.doi.org/10.58286/32432.