International Journal of Applied Resilience and Sustainability, Volume 2, Issue 2, March 2026, pp. 618-630

Published online at Deep Science

International Journal of Applied Resilience and Sustainability

Journal homepage: https://deepscipub.com/ijars

Artificial intelligence-driven intelligent concrete crack detection and real-time width estimation for smart structural health monitoring systems: A review

Mohammed Shakeebulla Khan [1], Adinath Rajendra Puri [2], Rosy Pradhan [3]

[1] *Department of Civil Engineering, St. John College of Engineering and Management (SJCEM), Palghar, Maharashtra, India*
[2] *Computer Engineering Marathwada Mitra Mandal's Polytechnic Pune*
[3] *Department of computer Engineering, St. John College of Engineering and Management (SJCEM), Palghar, Maharashtra, India*

Article Info:
Received 02 January 2026
Revised 27 February 2026
Accepted 02 March 2026
Published 05 March 2026

Corresponding Author:
Mohammed Shakeebullla Khan
E-mail: shakeeb.ulla@gmail.com

Abstract

It is partly the concrete infrastructure degradation (mainly crack formation and crack propagation) that is one of the most significant challenges facing the contemporary civil engineering. The traditional inspection procedures, which are based on a regular manual survey, non-destructive testing (NDT) guidelines, and an elementary image processing pipeline, have proven their constant incapacity to achieve sensitivity, scalability, and real-time reactivity. This combination of deep learning, computer vision, and embedded edge computing has also been the driving force behind a paradigmatic shift to the creation of intelligent, autonomous structural health monitoring (SHM) systems with the capability to operate in continuous, high-fidelity crack detection and quantitative width measurements. The present paper is a critical, interdisciplinary survey of the current state-of-the-art in AI-driven concrete crack detection that includes the following architectures, object detectors, semantic segmenters and detectors, and emergent paradigms of vision transformer (ViT) and vision-language paradigms (CLIP). Specific analytical focus is given to real-time crack width estimation techniques - such as pixel-to-millimetre calibration, stereo vision triangulation, laser profilometry, and Structure-from-Motion (SfM) 3D reconstruction techniques and their corresponding error metrics and deployment limits. The review also discusses how these AI components are integrated into smart SHM systems, which include edge AI systems, IoT data pipelines, cloud-edge hybrid systems, drone-based inspection systems, and digital twin systems. The critical research gaps that have been observed are lack of standardised benchmark datasets, lack of rigorously validated real time deployment studies, lack of cross-domain model generalisation, and unsolved hardware-software co-design issues.

Keywords: Structural health monitoring, Crack detection, Convolutional neural networks, Semantic segmentation, Deep learning, Vision transformers.

1. Introduction

1.1 Criticality of Crack Monitoring in Structural Health Monitoring

The concrete structures form the base of the world-built infrastructure including bridges, tunnels, dams, highways, high-rise and nuclear buildings and containment facilities. Mechanical degradation of these assets predominantly by crack initiation and propagation is a major failure mode of these assets that has extensive safety, economic, and societal consequences. Cracks are considered to be a first-hand indication of structural distress such as overloading, fatigue cycling, and reinforcement expansion due to corrosion, alkali-silica reaction, and settlement variation. Most importantly, the rich diagnostic data is encoded as crack width, depth, orientation, and propagation velocity and, when monitored continuously and with the appropriate degree of precision, allows undertaking the necessary actions before the occurrence of a disastrous failure. According to the estimates provided by American Society

https://doi.org/10.70593/deepsci.0202024

of Civil Engineers, the U.S. is facing a multi trillion dollar infrastructure repair gap with bridge decay being one of the most severe elements [1].

Devices that can automatically capture crack behavior, real-time crack warnings and generate useful maintenance information can be described as a paradigm shift capability in the management of infrastructure assets. The shift towards continuous and data-based condition monitoring as an alternative to the periodic inspection campaign is part of the larger trends of digitalization in construction and infrastructure where owners of assets are increasingly insisting on objective and quantitative condition information, as well as information that is dense in time and location, to facilitate risk-based maintenance scheduling and regulatory compliance.

1.2 Conditions of Manual and Conventional NDT

Visual inspection that is done manually is the most commonly employed crack detection method in the world. Structures are checked by trained inspectors who record the surface anomalies with crack gauges and photographs and prepare periodical reports of condition. Although this method has the advantage of human contextual judgement, it is essentially limited on the basis of subjectivity, varying as an inspector dependent variability, accessibility, sparsity in time, and failure to identify sub-millimetre cracks under varying lighting conditions or on a contaminated surface. Inter-rater reliability coefficients have always been reported to be less than 0.70 when assessing manual measurements of the crack width and this reduces the repeatability that is very necessary when tracking Easterlin over time [2].

Non-destructive testing (NDT) techniques, such as ultrasonic pulse velocity (UPV), ground-penetrating radar (GPR) and acoustic emission monitoring and impact-echo, have an extension of the inspection sensitivity into the sub-surface space, but introduce their limitations. UPV needs close contact with the transducer, and acoustic emission also has limitless scale. GPR requires expertise and post-processing skills. No of these modalities naturally yields the spatially continuous, picturally interpretable crack mapping most indicating of maintenance prioritisation.

1.3 The development of the Artificial Intelligence in infrastructure diagnostics has emerged

The coming of age of deep learning models and specifically convolutional neural networks in the wake of the historic ImageNet performance of Krizhevsky et al. (2012) [3] signaled a paradigm shift in the research of computer vision towards feature extraction guided by data. In the field of civil engineering, this change has allowed the substitution of manually designed image processing pipelines with trainable end-to-end models that can acquire discriminative representations directly trained on annotated crack imagery. These shifts in the state of the art of the field have been enabled by the presence of GPU-accelerated computing, open-source deep learning platforms (TensorFlow, PyTorch), and increasingly large matching annotated datasets, all of which have spurred the transition between field-deployable prototypes and their realization in the lab.

At the same time, with the advent of edge AI hardware platforms, such as NVIDIA Jetson boards, Google Coral TPUs, and Raspberry Pi microcomputers, on-device inference required by real-time SHM applications in infrastructure settings with either limited or intermittent connectivity has become a reality. The combination of AI-based crack detection with IoT sensor networks, cloud analytics platforms, and UAV inspection systems has increased the range of operation of intelligent SHM further.

1.4 The reason why Real-Time Crack Width Estimation is a motivated field is given in this section

Crack detection Binary or multi-classification of regions in images as either cracked or not detection is a required but not a sufficient requirement to operate SHM. Structural safety and maintenance scheduling engineering standards (ACI 224R, Eurocode 2, BS 8110) give precise limits on crack width limits above which remediation is required. In aggressive conditions with reinforced concrete, the width of surface cracks that are greater than 0.2-0.3 mm in width are linked to faster chloride penetration and reinforcement corrosion; in prestressed concrete, still smaller values are used. Therefore, proper quantitative width estimation is not only a research goal, but also an operational requirement.

2. Literature Survey: Traditional Crack Detection Techniques

2.1 Protocols to be followed in visual inspection

Normal visual inspection practices, which are documented in standards like the AASHTO Manual of Bridge Evaluation and the UK Highway Agency BD 63/07, involve regular close-range surveys by trained inspectors at given intervals - normally bi-annual inspection of bridges. The Crack mapping is done by hand drawing, photography, and direct measurement by using the crack comparator gauges, which are graded at 0.05 mm. Although such processes provide a baseline condition record, their inherent constraints, namely, subjective evaluation, observer exhaustion, impossibility of monitoring critical areas and failure to measure dynamic crack behavior, have created a longstanding interest in automated methods.

2.2 Ultrasonic and Acoustic NDT Methods

The ultrasonic pulse velocity (UPV) testing is used to measure the compressive wave transit times in concrete with the aberrant transit time showing the existence of crack planes or voids [4]. Stress wave events that are produced during propagation of a crack or reinforcement slip are recorded using acoustic emission (AE) monitoring, and may be used to detect active cracking passively. Time-of-flight diffraction (TOFD) and phased array ultrasonic testing (PAUT) provide better spatial resolution, but are very expensive and complex to operate, only used on critical infrastructure components.

2.3 Classical Image Processing Techniques

Before the deep-learning age, automated crack recognition had been based on classical image processing pipelines including image acquisition, pre-processing (contrast enhancement, noise reduction), feature extraction (thresholding, edge detection, morphological operations) and classification [5]. The thresholding of Otsu takes advantage of bimodal pixel intensity histograms to divide crack areas, and of the Canny edge detector to determine intensity gradients at the boundary of the cracks. These methods have acceptable performance in strongly controlled laboratory settings but perform badly when there is variation in illumination, surface heterogeneity, occlusion, and complicated crack morphologies typical of the infrastructure in the real world. The main weaknesses of classical image processing are that they rely on a manually tuned parameter, sensitivity to change in image qualities and lack of generalization when used with different types of infrastructure.

3. Methodology: Ai-Based Crack Detection Techniques

3.1 CNN-based Classification Models

Convolutional neural networks are trained in hierarchy learning to produce successive layers of convolution, batch normalization, non-linear activation, and pooling to produce spatial features representations. The deep homogeneous convolutional stacks of VGGNet [6] were one of the earliest architectures used to tackle the problem of crack image classification achieving accuracy over 98% on controlled benchmark datasets [7] with these stacks. ResNet [8] proposed residual skip connections which address the vanishing gradient issue in very deep networks, and allows one to use 50-152 layer models to extract more multi-scale features. EfficientNet [9] uses scaling of network width, depth and resolution of the network to compete with accuracy with reduced parameters as a result of which it is appealing to network edge deployment. MobileNet versions apply depthwise separable convolutions to minimize computational costs but still feature discrimination sufficiently to make use of the available features in a crack classification task.

3.2 Frameworks of Object Detection

Object detection models go further than binary classification to offer bounding box localisation. YOLO family works with the implementation of single-stage detection architectures, which can simultaneously estimate binding boxes and a probability of a class in a grid of anchor locations, and provide real-time inference rates that are needed to operate in the field. Kim and Cho [10] reported strong accuracy of encoder-decoder CNN-based crack detection on bridge images. Two-stage Faster R-CNN [11] uses region proposal network (RPN) to produce candidate crack regions then a classification and regression head, which usually has better localisation accuracy at a cost of higher inference latency.

3.3 Semantic Segmentation Models

Semantic segmentation models are utilized in classifying objects on the basis of the meaning or intention (Frisch 2010, p. 1).To map cracks at pixel scale, semantic segmentation models generate dense pixel-wise classification maps, in which the geometry of crack regions are directly coded. Originally created to perform biomedical image segmentation, U-Net [12] has been widely modified to crack segmentation. Encoder-decoder structure and skip connections between similar resolution levels allow it to localise the spatial information accurately and maintain the semantic information with IoU values above 0.80 on the conventional datasets of cracks [13]. DeepLabv3+ [14] uses atrous (dilated) convolutions and spatial pyramid pooling to learn multi-scale context, and can be seen to be scale-variation robust in terms of crack width

3.4 Vision Transformer Models

Vision Transformers (ViT) [15] break down input images into fixed-size patches sequences and use self-attention mechanisms to embed long-range spatial relationships. ViT-based models are especially favorable in detecting diffuse, branching crack networks in which the global context of structural features is diagnostically relevant in crack detection. Hierarchical versions like Swin Transformer [16] use shifted window attention and multi-scale extraction of feature learning which showed competitive results on dense prediction tasks like segmentation without compromising viable inference times.

3.5. Transfer Learning and Domain Adaptation

Using transfer learning (where models trained on large scale datasets (ImageNet, COCO)) are re-trained on crack-specific data has become the standard form of training crack detectors, due to the relative dearth of annotated infrastructure imagery. Interestingly, the few-shot learning methods [17] are specifically applicable to rare crack structures or new infrastructure varieties in which the labelled data are limited in reality. Domain adaptation methods deal with the distribution difference between source domains (e.g., bridge concrete) and the target domains (e.g., tunnel lining or pavement) with the help of adversarial training, style transfer, and data augmentation.

4. Results: Real-Time Crack Width Estimation

4.1 Image-Based Measurement Methodologies

Findings and estimations in 2D imagery of crack width are basically based on a form of matching the gap between pixels on a picture plane and dimensions on a physical object plane. The simplest method involves using reference objects with known size that are co-located in inspection images to estimate a pixel to millimetre scale factor [18]. Perpendicular distance transform techniques quantify crack width to be the shortest distance between crack opposing edges at a given point in the crack skeleton, giving spatially resolved width distributions. Measurement uncertainties of 0.0515mm in controlled conditions are attained by these approaches but when under perspective distortion, out of plane tilt and varying inspection distances, degradation occurs.

Table 1. Comparative analysis of AI-based crack detection models

Author(s)	Year	Model	Dataset	Acc/mAP	IoU/F1	Key Limitation
Zhang et al.	2016	VGG-16	Custom concrete (40k patches)	98.1%	—	Patch-level only; no width estimation
Zou et al.	2019	U-Net (DeepCrack)	DeepCrack (537 images)	—	IoU: 0.83	Under-segments hairline cracks
Kim & Cho	2018	Encoder–Decoder CNN	Custom bridge imagery	94.3%	—	Reduced accuracy on complex backgrounds
Dung & Anh	2019	ResNet-50 + UNet	Crack500	—	IoU: 0.86	GPU-dependent; not real-time on edge
Pauly et al.	2017	Faster R-CNN	Road pavement (2,000)	mAP: 84.7%	IoU: 0.78	Slow inference (~8 fps); inaccessible edge
Chen et al.	2021	DeepLabv3+	CrackForest + custom	97.2%	IoU: 0.81	Scale sensitivity in dense crack scenes
Fan et al.	2022	EfficientDet-D3	Tunnel lining dataset	mAP: 89.6%	F1: 0.88	Limited generalisation across infra. types
Hoskere et al.	2018	MobileNetV2 (Multi-scale)	UAV bridge imagery	94.8%	F1: 0.87	Reduced accuracy vs. larger CNNs
Liu et al.	2021	Swin Transformer	Mixed infrastructure	96.40%	F1: 0.91	High compute; not deployable on edge
Chen et al.	2023	CLIP fine-tuned	Custom multi-infra.	93.10%	F1: 0.85	Zero-shot; width estimation absent

Note: "—" denotes metric not reported. Acc = classification accuracy for patch-level models; mAP at IoU = 0.5 for detection models.

AI Model Performance: IoU vs. Real-Time Inference Speed for Crack Detection Architectures

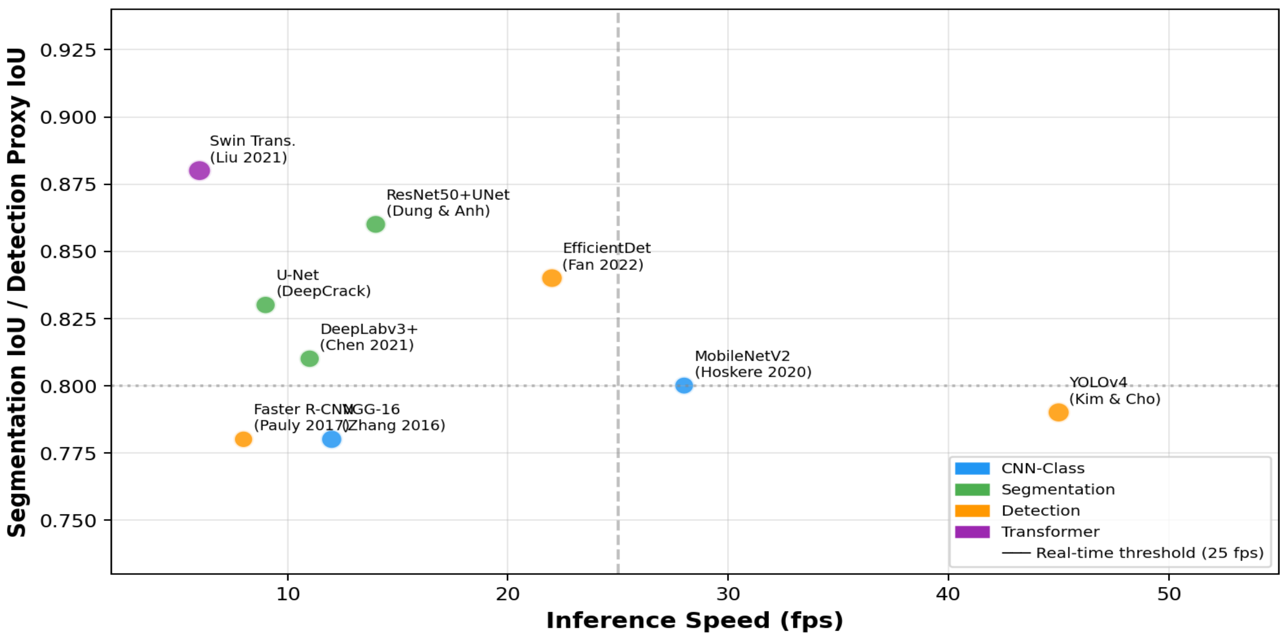

Figure 1. Comparative scatter plot of AI crack detection architectures by segmentation IoU versus real-time inference speed (fps). Bubble size indicates relative model complexity. The vertical dashed line marks the 25 fps real-time threshold; the

horizontal dashed line marks the IoU = 0.80 performance benchmark. Colour coding distinguishes architecture categories. Sources: compiled from references [7,10,13,14,16,26,27,28,29].

4.2 Pixel-to-Millimetre Calibration

Metrical conversion is sensitive and proper geometric camera calibration is necessary with either the planar calibration method (as presented by Zhang) [19] or the direct linear transformation (DLT) which takes into consideration the lens distortion parameters and projection geometry. In cases of mobile inspection systems - UAVs, ground robots, handheld devices - adaptive calibration with reference markers in sight simultaneously or Stereo baselines is required to consider the changeable camera-to-surface ranges. The analyses of uncertainty propagation show that a 10 percent error in distance measurement spreads to a similar relative error in width measurement, as the analysis shows the depth information is very essential.

4.3 Stereo Vision and 3D Reconstruction

Stereo vision systems utilise two calibrated cameras having known baseline to triangulate 3D crack geometry giving a depth resultant that facilitates measurement of width without dependence on distance. Multi-view stereo and Structure-from-Motion (SfM) on overlapping sequences of UAV photos re-create dense 3D point clouds of scanned surfaces, where the width of cracks can be measured in the physical coordinate system with uncertainty of less than 0.1 mm at inspection distances of less than 2 m [20]. Real-time 3D reconstruction is still a computationally expensive technique to work with embedded systems.

*4.4 IoT and Laser-Ba*sed Integration of M*easurements*

In laser triangulation systems, the pattern of lines or dots on the inspection surface is provided with a laser and geometric distortion of the pattern is measured using a camera positioned laterally displaced to obtain a topography of the surface with a resolution in the micrometre scale [22]. IoT-based crack gauges based on resistive, capacitive, or fibre Bragg grating (FBG) sensing technologies can use continue monitoring of the crack opening displacement at discrete points in space, to complement AI-based spatial mapping with high-temporal-resolution point measurements.

*4.5 Metrics and Measure*ment Issues o*f error*

The three most commonly used quantitative measurements of the crack width estimation performance include Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and relative percentage error based on measurements on ground truth of the optical microscopy or contact gauge measurements of high preciseness. According to the literature, MAE values are 0.02 mm in case of laboratory stereo vision devices [21] and 0.25 mm in case of hand-held smartphone-based systems in outdoor conditions, respectively, which are a considerable difference between controlled and operational conditions.

The main issues facing the real-time estimation of crack widths are: the change in illumination to give inconsistent contrast at the crack boundary; the noise in surface texture creating artificial crack boundaries; the overwhelming class disparities between the crack and non-crack pixels making it difficult to train; and the inability to see hairline cracks (below 0.1 mm in width) which are diagnostically significant but close to the optical resolution of the standard inspection cameras.

Table 2. Crack width estimation methods-performance comparison

Author(s)	Year	Method	MAE (mm)	RMSE (mm)	Key Finding
Xu et al.	2019	Stereo vision + CNN	0.020	0.031	Sub-0.1 mm accuracy in lab; degrades >1.5 m
Adhikari et al.	2014	Calibrated monocular + ResNet	0.080	0.120	Practical for field use with reference calibration
Mohan & Poobal	2018	Smartphone camera + IP	0.250	0.380	Low-cost but insufficient for structural assessment
Yang et al.	2022	Laser profilometer + YOLO	0.015	0.022	High accuracy; limited to accessible surfaces
Fan et al.	2021	UAV SfM 3D reconstruction	0.060	0.090	Distance-independent; scalable to large structures

Note: IP = image processing; SfM = Structure-from-Motion. Ground truth via calibrated optical microscopy or contact gauge in all studies.

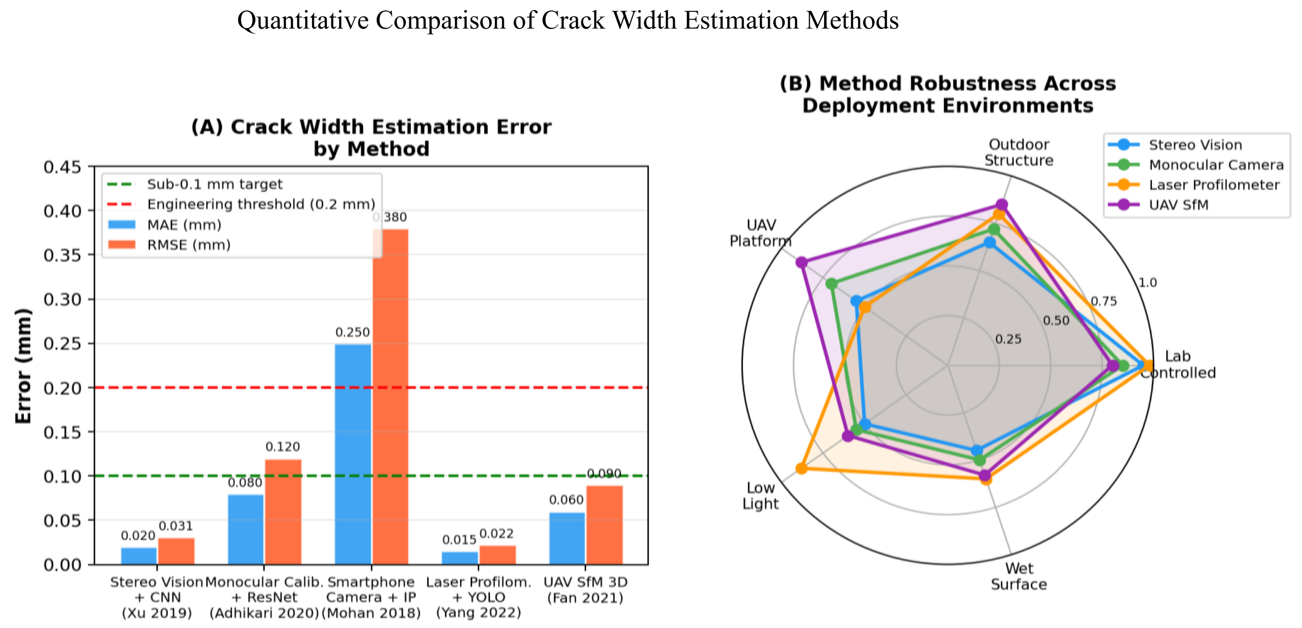

Figure 2. (A) Bar chart comparison of Mean Absolute Error (MAE) and RMSE for five crack width estimation methods. The green dashed line marks the sub-0.1 mm engineering target; the red dashed line marks the 0.2 mm intervention threshold mandated by ACI 224R and Eurocode 2. (B) Radar plot of robustness across five deployment environments. Sources: [21,18,5,22,26].

5. Smart Structural Health Monitoring Systems

5.1 Conceptual Architecture of an AI-based SHM Device.

A full AI-based SHM system addressing concrete crack monitoring consists of 5 functionally different layers including (1) data collection, (2) on-device inference, (3) local data management, (4) network communication, and (5) cloud-based analytics. In the acquisition camera, inspection images are captured by a multi-spectral (or high-resolution) visible-light camera module at the acquisition layer (optionally with a laser profilometer and environmental sensors). Depending on its deployment requirements, this unit can be attached as a fixed node onto a bridge deck, fitted on a robotic crawler, or suspended under a UAV inspection platform.

The inference engine is located on an embedded AI accelerator - NVIDIA Jetson Nano, Google Coral Edge TPU or similar - and is running an optimised crack detection and segmentation model (usually at INT8 (quantised) precision to achieve throughput). Communication layer The layer sends summarised condition data over cellular (LTE/5G), Wi-Fi, or LoRaWAN connections to a cloud analytics platform. The cloud layer combines information of various monitoring nodes and communicates with BIM and digital twin models [23].

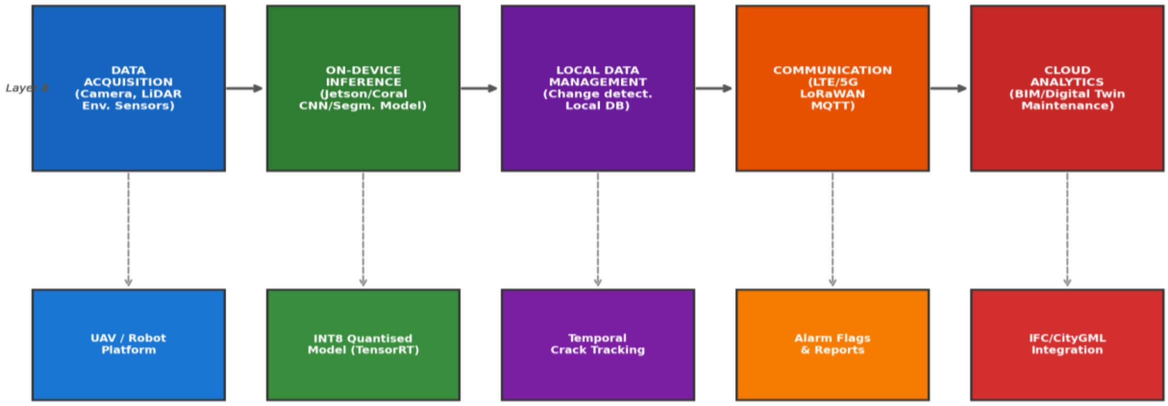

Figure 3. Five-layer smart SHM system architecture for AI-based concrete crack monitoring. Each layer performs a distinct function: data acquisition (Layer 1), on-device AI inference (Layer 2), local data management (Layer 3), network communication (Layer 4), and cloud analytics with digital twin/BIM integration (Layer 5). Sub-components for each layer are shown below.

5.2 Edge AI Systems and Inference on a chip

NVIDIA Jetson Nano offers 472 GFLOPS of GPU computing performance, which consumes 5 10 W of power, allowing the inference of small segmentation models (MobileNetV2-UNet) at 1525 fps on 640 480 pixel input images. This is extended to around 275 TOPS of tensor-optimised compute on the Jetson AGX Orin. The GOogles Coral TPU modules are characterized by an incredible energy saving in INT8 inference of lightweight models. Edge deployment Model optimisation uses pruning, knowledge distillation and after-training quantisation, and published works have shown 36x throughput and 24x memory savings with INT8 quantisation, and 24 accuracy reduction by absolute IoU [23].

5.3 IoT Implementation and Pipelines

SHM IoT implementations scatter smart sensor nodes throughout the extensive structural structures such as bridge spans, tunnel rings, building facades, and each sensor node does local AI inferences and sends condition summaries over standardised protocols (MQTT, CoAP, HTTP REST). Scalable, message brokering, time-series storage, and rule-based alerting are offered in data ingestion platforms (AWS IoT Core, Azure IoT Hub, ThingsBoard). The low-latency edge inference to detect real-time anomalies and cloud-based batch analytics to data to determine longitudinal trends is the best architectural pattern to adopt when deploying SHM on a large scale.

5.4 UAV-Based Crack Inspection

Unmanned aerial vehicles have come in as exceptionally well-performing systems to crack surveying of large, geometrically intricate, or inaccessible structures such as bridge soffits, dam faces, tall building fronts and wind turbine towers [20]. UAVs used in inspection have high-resolution camera pay loads, usually complemented with LiDAR or structured light sensors, and operate under pre-programmed flying profiles at predictable standoff ranges to ensure repeatable image scale. Onboard AI inference allows real-time crack sensing and GPS-referenced mapping of the aircraft in flight and post-flight SfM reconstruction creates centimetre-resolution 3D models to allow correct width measurement. Rigid regulatory requirements, wind sensitivity, battery duration (usually 2030 minutes) are all major operational issues.

6. Multimodal and Advanced AI Approaches

6.1 CLIP and Vision-Language Models

Contrastive Language-Image Pre-training (CLIP) and later vision-language foundation models are trained to acquire similar visual and textual representations based on enormous web-scale data sets enabling zero-shot and few-shot inference on new visual tasks via natural language prompting. Early research [24] indicates that fined tuned CLIP variants can obtain competitive crack classification behavior using minimal labelled data implying a positive outlook with respect to generalisation between the typologies of infrastructure without retraining per-domain.

6.2 Few-Shot Learning and Self-Supervised Learning

Few-shot learning models allow effective crack detection with as little as five to twenty labelled examples per class, which is a solution to the endemic lack of data of many specialised interfaces in infrastructure inspection. Designed Prototypical Networks, MAML and Matching Networks have been scaled to crack classification and detection problems and have shown ability to quickly adapt to new types of cracks [17]. Self-supervised pre-training approaches SimCLR (Masked contrastive learning), MoCo, MAE (masked image modelling), BEiT: Unlabelled inspection imagery is used to learn powerful visual representations.

6.3 Digital Twin Integration

Dynamic computational models co-ordinated with physical structures via data provided by continuous sensors (digital twins) give the semantic information required to transform AI detection of cracks on images into an integrated assessment of the condition of the structure as a whole. Crack maps found by AI can be automatically mapped to the geometric model of the digital twin so that spatially indexed crack inventories can be stored between inspection campaigns. When BIM environments (Autodesk Revit, Bentley OpenBuildings) are integrated with AI crack monitoring data using standardised data exchange formats (IFC, CityGML), then field condition data that is linked to the maintenance scheduling process and lifecycle cost analysis can be used.

7. Discussion

7.1 Performance Assessment Model

Comparable AI crack detection methods cannot be made without the prior performance evaluation. In the case of binary crack classification, the standard set is Accuracy, Precision, Recall, and F1-score; in crack detection, where the cost of a false negative (missed crack) is much higher than the cost of a false positive, Recall is usually the most important performance measure. In the case of object detection models, localisation and classification performance are combined by the mean Average Precision (mAP) as functions of the IoU thresholds. Measurement accuracy of crack width is measured using MAE, RMSE, and relative percentage error and Bland-Altman analysis gives agreement limits of automated and reference measurements.

7.2 Comparative Analysis and Critical Observations

A comparative analysis conducted in Tables 1 and 2 indicates a number of important observations. First, there is no AI architecture that can be used in all practically relevant deployment situations, segmentation models (U-Net, DeepLabv3+) can provide pixel-level granularity required to measure width, but require intensive computational resources; object detectors can perform real-time inferences at lower geometric accuracy; transformer models can provide better global context modelling, but at a high computational cost. Second, the relation between benchmark dataset performance and operational field accuracy is not well defined in the literature and most studies do not have formal measures of uncertainty or field validation guidelines. Third, performance based on crack width estimation would significantly worsen between laboratory and field conditions. The MAE(smartphone based) (0.25mm) is greater than 0.2mm engineering limit that would initiate maintenance action in most structural standards and this indicates the inefficiency of the existing field-ready width estimation techniques.

The numerical under-representation of quantitative width estimation in the AI crack detection literature is both an indication of the increased technical challenge of the measurement problem, and a history of segregation between computer vision scholars and structural metrologists. The solution to this disconnect is to incorporate metrological rigour into the processes of AI development, by developing interdisciplinary research programmes that integrate metrological rigour into the process of finding solutions.

7.3 Research Gaps Structured Analysis

The five identified gaps in Table 3 are not independent of each other - they multiply each other in a manner that seriously slows the pace of the deployment of operational SHM. The lack of standardised datasets does not allow rigorous cross-method comparison, field validation protocols do not ensure improvements on the apparent benchmark do not translate into operational improvements, and problems in hardware-software co-design do not allow the efficient deployment of a model on the edge platforms needed to make SHM practical.

Table 3. Structured analysis of key research gaps, current state, and proposed directions

Research Gap	Current State	Proposed Direction
Standardised benchmark datasets	Fragmented datasets (CRACK500, DeepCrack, CrackForest) inconsistent annotation, limited infrastructure diversity	Large-scale, multi-infrastructure, metrologically validated open benchmark
Real-time deployment validation	Lab/controlled-condition evaluations dominate; operational field validation is rare	End-to-end field studies with metrological rigour and statistical uncertainty reporting
Cross-domain generalisation	Models trained on pavement fail on bridges/tunnels without fine-tuning	Domain adaptation, meta-learning, and domain-randomised augmentation frameworks
Hardware-software co-design	Model accuracy and edge compute constraints evaluated independently	Co-optimised NAS for target hardware; standardised SHM AI benchmarking protocols
Longitudinal width estimation at scale	Width estimation confined to static lab setups; no long-term SHM studies	UAV-adaptive calibration; temporally consistent crack tracking with uncertainty bounds

Note: NAS = Neural Architecture Search; infra. = infrastructure.

8. Future Research Directions

8.1image TinyML and Extreme Edge Deployment

TinyML involves optimisation and execution of machine learning models in microcontroller-sized (ARM Cortex-M, RISC-V) platforms (under milliwatt power constraints) to allow literally ubiquitous battery-powered nodes that monitor cracks. Microcontroller-optimised neural architecture search (NAS) algorithms can optimally identify CNNs with competitively high accuracy of crack detection with memory footprints of just a few kilobytes. TinyML frameworks (TensorFlow Lite Micro, CMSIS-NN, Edge Impulse) are quickly growing the capabilities of exceptionally resource-constrained platforms - although themetrological estimated width estimation in TinyML conditions is still an open challenge on par.

8.2 Centre of Autonomous Inspection Robotics

AI-powered inspection robots on the ground offer complementary benefits to UAVs - more payload capacity due to high-endurance sensor payloads (laser profilometers, LiDAR, acoustic sensors), they can be operated indefinitely with no battery charges, and can be used to inspect areas that are closed like tunnel interiors or building basements. Legged robots (Boston Dynamics Spot, ANYbotics ANYmal) offer agile movement on rugged surfaces. The integration of real-time crack detection with autonomous path planning, i.e., an adaptive approach in which the robot trajectory is steered toward areas of high crack density, is an interesting feature towards effective inspection of large scale.

8.3 Self-learned and Never-ending Learning SHM Systems

Self-learning SHM systems use the uninterrupted information provided by operation monitoring to optimize the model of crack detection with time without manual annotation or retraining. Active learning models find the most informative images in the monitoring data stream that can be annotated by experts and reduce the labelling work, maximizing the model enhancement rate. Continual learning algorithms online update weights of the models on a continual basis using new observations and avoiding catastrophic forgetting.

8.4 BIM, Digital twins and multi-modal sensor Fusion

The most significant immediate effect of AI-crack detection is the interaction with the BIM and digital twin systems that would relate the data on the field condition to the engineering analysis, lifecycle management, and regulatory reporting processes. To achieve interoperability Standardised data description of crack observations - geometry, width, depth, location, confidence, and measurement provenance Standardised data descriptions based on IFC and the built environment ontologies should be created. The next generation SHM system will add visual crack identification and acoustic visualisation of the localisation of acoustic emission, distributed fibre optic, and sub-surface GPR AI analysis.

9. Conclusion

This review has critically analyzed the state-of-the-art in AI-assisted concrete crack detection and real-time width prediction in smart structural health monitoring, the range of technical structures beginning with the classical CNN architectures, a range of new vision transformer and vision-language frameworks, and the range of laboratory measurement systems to deployed operational edge AI systems. The synthesis demonstrates a field of dynamic transition: on the one hand, technically, it has reached its peak with the laboratory-proven detection capabilities, and on the other hand, it remains in many ways undeveloped in the field of metrological width estimation at scale, cross-domain generalisation, as well as operated deployment.

This review has undertaken a critical examination of the state-of-the-art in AI-driven concrete crack detection and real-time width estimation for smart structural health monitoring, spanning the technical landscape from foundational CNN architectures to emerging vision transformer and vision-language paradigms, and from laboratory measurement methodologies to operational edge AI deployment systems. The synthesis reveals a field in dynamic transition: technically mature in laboratory-validated detection capabilities, but still substantially immature in the dimensions of metrological width estimation at scale, cross-domain generalization, and rigorously validated operational deployment.

The comparative analysis of AI model architectures demonstrates that no single approach dominates across all practically relevant deployment scenarios. Segmentation models offer pixel-level granularity essential for width measurement but challenge edge deployment; object detectors enable real-time inference at the cost of reduced geometric precision; and transformer models offer superior global context modelling at elevated computational cost. Transfer learning from large-scale pre-trained models remains the most practically effective training strategy for data-sparse applications, while domain adaptation techniques require further development to achieve reliable cross-infrastructure generalization.

Crack width estimation emerges from this review as the most consequential and least resolved frontier in AI-based SHM. Engineering standards mandate width thresholds requiring sub-0.1 mm field accuracy, a target current AI systems approach only in controlled laboratory scenarios. Closing the gap between laboratory-demonstrated accuracy and field-validated precision requires systematic research into calibration stability, illumination robustness, and uncertainty quantification currently underrepresented in the literature. At the system integration level, the convergence of edge AI acceleration hardware, IoT data infrastructure, UAV platforms, and digital twin architectures has created the technical substrate for transformative SHM capabilities but the interdisciplinary gap between AI researchers and civil infrastructure practitioners continues to retard translation from research prototype to operational deployment. Sustained collaborative engagement between computer vision, embedded systems, structural engineering, and metrology communities, alongside investment in standardized benchmark datasets and field validation protocols, is the critical enabling condition for realizing the transformative potential of AI-enabled structural health monitoring.

Author Contributions

MSK: Conceptualization, study design, analysis, data collection, methodology, software, resources, visualization, writing original draft, writing review and editing, and supervision. ARP: Conceptualization, study design, analysis, data collection, methodology, software, resources, visualization, writing original draft, writing review and editing, and supervision. RP: Conceptualization, study design, analysis, data collection, methodology, software, resources, visualization, writing original draft, writing review and editing, and supervision.

Conflict of interest

The authors declare no conflicts of interest.

References

[1] Yingqun Z, Zhuang L, Nana Z. Artificial Intelligence-Driven Health Monitoring and Early Warning Technology for Concrete Structures. International Journal of High Speed Electronics and Systems. 2026 Jun 28;35(03):2540450. https://doi.org/10.1142/S0129156425404504

[2] Vairagade VS. Artificial intelligence-driven predictive modeling of multi-functional carbon nanotube infused smart cement for structural reinforcement and real-time damage sensing. Frontiers of Structural and Civil Engineering. 2025 Sep;19(9):1403-17. https://doi.org/10.1007/s11709-025-1219-y

[3] Lazim TM, Mohd NA. AI-Driven Structural Health Monitoring and Seismic Safety for Next-Generation Urban Infrastructure.

[4] Sarma IV, Chanda S, Reddy MS. The convergence of artificial intelligence and structural health monitoring: a comprehensive review of methodologies, advancements, and future trajectories. Life Cycle Reliability and Safety Engineering. 2025 Dec;14(4):531-43. https://doi.org/10.1007/s41872-025-00364-z

[5] Li F, Luo D, Niu D. Data-intelligence driven methods for durability, damage diagnosis and performance prediction of concrete structures. Communications Engineering. 2025 Jun 3;4(1):100. https://doi.org/10.1038/s44172-025-00431-4

[6] Farrar CR, Worden K. An introduction to structural health monitoring. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2007 Feb 15;365(1851):303-15.

[7] Shahin M, Chen FF, Maghanaki M, Hosseinzadeh A, Zand N, Khodadadi Koodiani H. Improving the concrete crack detection process via a hybrid visual transformer algorithm. Sensors. 2024 May 20;24(10):3247. https://doi.org/10.3390/s24103247

[8] Masalila S, Sheikh S, Firoozi A. Toward Resilient and Intelligent Infrastructure: A Survey of Smart Monitoring Systems and Emerging Technologies. Available at SSRN 5353106. 2025. https://doi.org/10.2139/ssrn.5353106

[9] Li D, Li B, Wang T, Zhang J. Lightweight structural health monitoring and safety evaluation: Review and case studies. Advances in Structural Engineering. 2025 Sep;28(12):2157-79. https://doi.org/10.1177/13694332251325043
[10] Islam MM. Machine learning-based predictive modeling for assessing bridge load capacity using real-time sensor data. Journal of Sustainable Development and Policy. 2025 Sep 30;4(03):01-37. https://doi.org/10.63125/v5y21788
[11] Silva-Campillo A, Herreros-Sierra MA, Pérez-Arribas F. Acoustics as a Structural Health Monitoring Tool in Naval and Offshore Structures: A Comprehensive Review. Applied Sciences. 2026 Feb 2;16(3):1477. https://doi.org/10.3390/app16031477
[12] Yamsani N, Kaur A, Manwal M, Bansal G. CrackNet: A CNN-based Framework for Automated Concrete Crack Detection. In2025 6th International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS) 2025 Dec 15 (pp. 2048-2053). IEEE. https://doi.org/10.1109/ICICNIS66685.2025.11315680
[13] Živković M, Žujović M, Milošević J. Architectural 3D-Printed structures created using artificial intelligence: A review of techniques and applications. Applied Sciences. 2023 Sep 26;13(19):10671. https://doi.org/10.3390/app131910671
[14] Abbas SR, Seol H, Abbas Z, Lee SW. Exploring the role of artificial intelligence in smart healthcare: a capability and function-oriented review. InHealthcare 2025 Jul 8 (Vol. 13, No. 14, p. 1642). MDPI. https://doi.org/10.3390/healthcare13141642
[15] Wahab Sait AR, Sankaranarayanan S, Yu Y. Vision transformers-Kolmogorov-Arnold networks-based consumer driven surface cracks classification model. Scientific Reports. 2026 Feb 15. https://doi.org/10.1038/s41598-026-40359-z
[16] Gill S. Artificial Intelligence in Concrete Crack Detection and Structural Health Monitoring.
[17] Prakash V, Debono CJ, Musarat MA, Borg RP, Seychell D, Ding W, Shu J. Structural health monitoring of concrete bridges through artificial intelligence: A narrative review. Applied Sciences. 2025 Apr 27;15(9):4855. https://doi.org/10.3390/app15094855
[18] Sofi A, Regita JJ, Rane B, Lau HH. Structural health monitoring using wireless smart sensor network-An overview. Mechanical Systems and Signal Processing. 2022 Jan 15;163:108113. https://doi.org/10.1016/j.ymssp.2021.108113
[19] Azanaw GM. INTEGRATING AI IN STRUCTURAL HEALTH MONITORING (SHM): A SYSTEMATIC REVIEW ON ADVANCES, CHALLENGES, AND FUTURE DIRECTIONS. i-Manager's Journal on Structural Engineering. 2024 Dec 1;13(3). https://doi.org/10.26634/jste.13.3.21791
[20] Ghosh A, Edwards DJ, Hosseini MR, Al-Ameri R, Abawajy J, Thwala WD. Real-time structural health monitoring for concrete beams: A cost-effective 'Industry 4.0'solution using piezo sensors. International Journal of Building Pathology and Adaptation. 2021 Mar 11;39(2):283-311. https://doi.org/10.1108/IJBPA-12-2019-0111
[21] Khan MZ, Shahzadi M, Khan A, Ali U, Hassan MA, Hussain M. Review on crack detection in civil infrastructure using structural health monitoring and machine learning techniques. Innovative Infrastructure Solutions. 2025 Aug;10(8):348. https://doi.org/10.1007/s41062-025-02147-y
[22] Sohn H, Farrar CR, Hemez FM, Shunk DD, Stinemates DW, Nadler BR, Czarnecki JJ. A review of structural health monitoring literature: 1996–2001. Los Alamos National Laboratory, USA. 2003 Feb;1(16):10-2989.
[23] Birgani SA, Zadeh SS, Davari DD, Ostovar A. Deep learning applications for analysing concrete surface cracks. International Journal of Applied Data Science in Engineering and Health. 2024 Oct 26;1(2):69-84.
[24] Hossain MI. Deployment of AI-supported structural health monitoring systems for in-service bridges using IoT sensor networks. Journal of Sustainable Development and Policy. 2022 Dec 1;1(04):01-30. https://doi.org/10.63125/j3sadb56
[25] Yingqun Z, Zhuang L, Nana Z. Artificial Intelligence-Driven Health Monitoring and Early Warning Technology for Concrete Structures. International Journal of High Speed Electronics and Systems. 2026 Jun 28;35(03):2540450. https://doi.org/10.1142/S0129156425404504
[26] Sony S, Laventure S, Sadhu A. A literature review of next-generation smart sensing technology in structural health monitoring. Structural Control and Health Monitoring. 2019 Mar;26(3):e2321. https://doi.org/10.1002/stc.2321
[27] Zinno R, Haghshenas SS, Guido G, VItale A. Artificial intelligence and structural health monitoring of bridges: A review of the state-of-the-art. IEEE access. 2022 Aug 17;10:88058-78. https://doi.org/10.1109/ACCESS.2022.3199443
[28] Mardanshahi A, Sreekumar A, Yang X, Barman SK, Chronopoulos D. Sensing techniques for structural health monitoring: A state-of-the-art review on performance criteria and new-generation technologies. Sensors. 2025 Feb 26;25(5):1424. https://doi.org/10.3390/s25051424
[29] Barbhuiya S, Das BB. Artificial Intelligence in Damage Detection of Concrete Structures: Techniques, Integration and Future Directions. InDamage Detection and Structural Health Monitoring of Concrete and Masonry Structures: Novel Techniques and Applications 2025 Mar 22 (pp. 31-92). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-97-8975-7_2