ICMESR 2026, India, 04-05 May 2026
https://doi.org/10.55084/gcp/001401

Intelligent Deep Learning Framework for Accurate Crack Detection and Width Measurement

Kanchan Dhapekar[1*] and Divya Prakash[2]

[1]Department of Civil Engineering, Poornima University, Jaipur Jaipur, Rajasthan.
[2]Department of Civil Engineering & Centre of Excellence in Water and Clean Air, Poornima University, Jaipur Jaipur, Rajasthan, India

Email: kanchan1983@gmail.com ([*]Corresponding Author)

Received : 17 May 2026; Accepted : 14 Jul 2026; Published : 08 Aug 2026

Abstract. The automated crack detection and width measurements in concrete structures are important to structural health monitoring, but the manual inspection of crack is inconsistent, and the traditional image processing methods are difficult to adapt under different illumination environments and conditions. In this study, a novel hybrid deep learning network is designed using the convolutional neural network and transformer networks to effectively identify and segment cracks on complex concrete surfaces under different scenarios. For segmented images, the procedure of skeletonization returns the center of the crack for the whole image allowing measurements of the crack width at multiple positions across the width of the crack that are perpendicular to the centerline. An FOV based calibration step follows, which translates the measurements as obtained from the pixels to accurate real-world metric measurements. Its combination of deep-learning, skeleton extraction and calibration allows for fully-automated quantitative crack characterization. Experimental results are shown with high accuracy and reliability in crack detection, segmentation and width estimation compared with classic techniques which suggest the capability for real-time and automatic inspection of infrastructures through automation provided by CSE.

Keywords: *Crack Detection, Crack Width Measurement, Deep Learning, CNN–Transformer, Structural Health Monitoring.*

1. INTRODUCTION

Concrete usage is very prevalent in the civil engineering industries such as bridges, buildings, tunnels, pavements, dams, retaining walls, etc. because of the strength and durable properties and also due to its economic aspects. However, these structures are constantly subjected to the environment, cyclical loading, temperature, shrinkage, corrosion and ageing which affect them and often cause problems of cracking [1],[2]. Any crack in any concrete or masonry structure is not simply a bad look and it serves as a good tool for not only assessing problems but also the safety of the building. Thus, good quality crack detection and quantification play an integral role in the integrity assessment, serviceability assessment and longevity of infrastructure systems [1], [2], [3].
Generally, the inspection of cracks is carried out by manually visual examination of the crack by engineers/inspectors. Manual inspection is quite common, but there are many drawbacks to "manually inspect", such as subjectivity, complexity, high usage of manpower in inspection or sometimes a high cost of the inspection. Furthermore, it is difficult and even risky to physically inspect various locations which are not easily reached or unsafe, such as tunnels, bridge tops and high-rises [4],[5].

© 2026.This is an open access article and allows to use the as per CC BY-NC-SA 4.0 license
ISBN: 978-81-996627-3-5

Traditional image processing operations like thresholding, edge detection and morphological operations are also applied to detect the cracks as an additional application. However, they are also sensitive to levels of noise, illumination differences, shadows, texture variations, and surface complexity, limiting their performance in real-world environments [4],[6],[7].

AI and deep learning technologies have greatly boosted the performance of vision-based automated inspection for structural health monitoring [8], [9], [10]. The ability of deep learning models, especially Convolutional Neural Networks (CNNs), has recently been unmatched in the field of image classification, object detection, and semantic segmentation [8], [11], [12]. CNNs are very useful for learning local spatial features such as edges, crack patterns, and texture details [11], [12], [13].

However, CNN-based models may not adequately address long-range dependencies and global contextual information, particularly in complex crack images containing multiple backgrounds and orientations [14], [15]. Recently, self-attention-based Transformer networks have demonstrated excellent capability in capturing global contextual information and long-range feature relationships, leading to improved performance in computer vision applications [9][15]. The combination of CNNs and Transformer networks is therefore an effective approach that leverages the advantages of both architectures for improved crack detection and segmentation performance [9][15].

In this study, Roboflow Detection Transformer (RF-DETR) a hybrid CNN–Transformer framework is proposed for crack detection and accurate crack width measurement using a hybrid CNN–Transformer architecture. The proposed framework combines advanced deep learning techniques for crack segmentation with skeletonization and Field-of-View (FOV)-based calibration methods to achieve accurate crack quantification [7], [15], [16]. In the first stage, the hybrid model processes concrete surface images to accurately identify and segment crack regions. Subsequently, crack centrelines are extracted using skeletonization to preserve crack geometry for further analysis [15], [16]. Crack width measurements are performed at multiple locations perpendicular to the extracted skeleton path and converted into real-world metric dimensions through FOV-based calibration [15], [16].

The major contributions of this study are summarized as follows:

1. Development of a robust hybrid CNN–Transformer framework for crack detection and segmentation [9], [15].
2. Crack center lines extraction using skeletonization technique and geometric analysis [15], [16]. FOV based calibration of pixel to real world size [15].
3. Automated measurement of crack width along a crack path at several locations.
4. Experiments performed for different lighting and surface conditions.

This proposed framework offers an efficient, accurate and intelligent solution for automated inspection of infrastructure and real-time, automated structural health monitoring applications [1], [9], [17].

2. LITERATURE REVIEW

2.1 Traditional Image Processing and Machine-Learning-Based Methods

The first automation solutions for the detection of cracks used the classic methods of image processing including thresholding, detection of edges, growth and morphological operations of regions. They showed high sensitivities to background scenes, surface texture, varying illumination and were not very reliable computational strategies. Later on, machine learning methods using handcrafted feature extraction methods were applied to further enhance the detection accuracy, namely Support Vector Machines (SVM), Random Forests (RF) and Artificial Neural Networks (ANN) [12]. The benchmark datasets that are publicly available (e.g., SDNET2018 [18]) have also been useful for early modelling and evaluation of crack-classification models.

2.2 CNN-Based Crack Detection Methods

Initial automation of crack detection used traditional image processing algorithms like thresholding, edge detection, region growing and morphological operations [19]. These methods were time efficient, but were not very robust as they were sensitive to changing lighting conditions, more complex background scenes, and different surface texture conditions [6], [20]. The detection accuracy has been enhanced by machine learning techniques with handcrafted feature extraction, such as Support Vector Machines (SVM), Random Forests (RF) and Artificial Neural Networks (ANN) [12].

With the introduction of Convolutional Neural Networks (CNNs) automatic feature learning contributed a further enhancement to the field of crack detection [13], [19]. The use of CNN-based models like U-Net, Faster R-CNN, Mask R-CNN, and even YOLO has been adopted to detect, localize, and segment cracks [6], [14], [21] [22]. Ali et al. [24], and Hacıefendioğlu and Başağa [14] proved the efficiency of CNNs in identification of the crack position accurately under various conditions, while Yang et al. [22] proved it as one of the most efficient methods for accurate crack identification in extreme environments.

The CNN-based approach has also been used for the crack width measurement by skeletonization, contour analysis, and distance transform technique. The FOV-based calibrations are prevalent and often employed for converting the measured crack dimensions in pixels to physical dimensions of cracks, which simultaneously have the advantage of low complexity and high accuracy [9], [21], [23].

2.3 Transformer-Based Crack Detection Methods

Recent work on deep learning has turned to Transformer-based networks and converged hybrid CNN–Transformer networks that are able to capture long-range (extra-pixel) spatial relationships in the network using self-attention mechanisms [3][4] that greatly surpass the receptive field of classical CNNs. Prominent representation of general features at a global level and good performance in more difficult tasks such as crack segmentation and detection of structural defects are strength points of these models.

Local feature extraction by CNN and global contextual information by Transformer have demonstrated good results in the quantification and precise segmentation [2], [17] under challenging environmental conditions. Egodawela et al. [2] developed an aerial crack detection model based on deep learning with Unmanned Aerial Vehicle (UAVs) and drew a conclusion that the Transformer-based approach worked well in the detection of cracks on large scale infrastructures. Similarly, considering the applicability of the advanced deep learning models in the field of structural health monitoring, Cha et al. [17] explained that models based on the Transformer architecture have come into incredible importance recently owing to their ability to detect the structural damage with high accuracy and support in developing smart and intelligent system management techniques for inspecting infrastructure [22].

After adopting the Transformer-based model, the crack detection and segmentation capability has been greatly improved, but the width of the cracks is still hard to calculate. In recent studies, the geometric image processing and reliable calibration methods (like Field-of-View [FOV] based calibration) are combined with the Transformer based detection frameworks to arrive at precise quantitative crack measurements that can be used in SHM applications, which is a major challenge [21].

3. PROPOSED METHODOLOGIES

In order to accurately recognize and segment cracks, CNN, which has the ability to learn local features, and Transformer, which has the ability to learn the global context, are selected as the learning structure for this work [2], [17]. Common image processing principles are followed for pre-processing to address common problems like noise, contrast, data normalization, data augmentation to make the model more robust [2], [11]. In the present study, the geometry of the crack is obtained by cracking image analysis in previous crack quantification studies [21],[23] and is used to perform a skeletonization of the cracks in the FOV-based calibration model.

However, the simulation shows the proposed RF-DETR model is well performing in accordance with the recent advances of deep learning in structural health monitoring [2],[6],[24]. These results again demonstrate the applicability of a hybrid approach to real-world systems and validate the use of a reliable and automatic solution to detect and measure cracks. The work provides good room for the development of intelligent systems for SHM, but it would be beneficial for future research to develop systems with more advanced real-time technology for infrastructure.

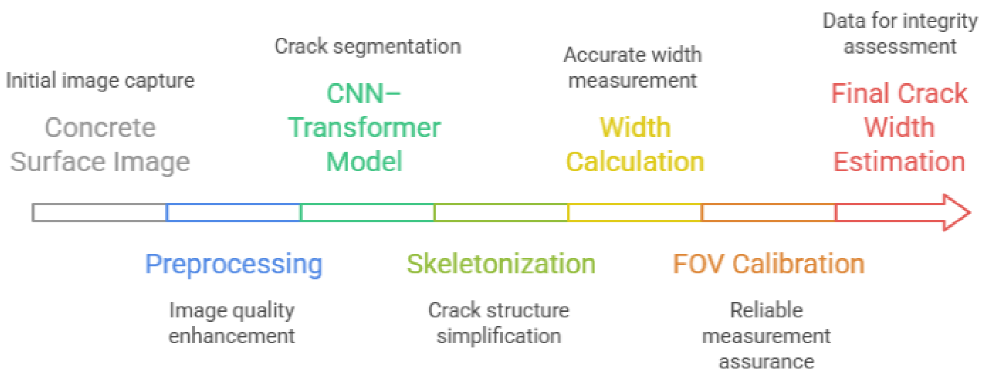


Fig. 1 Proposed Methodology

The overall work flow of the proposed crack width measurement framework is shown in Figure 1. The process starts from capturing image from the concrete surface and then processes the acquired image to improve the image quality in order to obtain the best results. After pre-processing the image, the obtained image is further analyzed using CNN–Transformer model for crack segmentation. Following skeletonization of the crack structure, the crack width, and the field of view (FOV) calibration is performed for an accurate measurement. Lastly, estimates of crack width are produced for integrity check of structures.

3.1 Image Acquisition and Preprocessing

A high-resolution digital camera is used to take a series of images of the concrete surface under a wide range of ambient lighting. Images acquired are scaled to a fixed resolution for both learning as well as inference purposes. It aims to preprocess images to better improve the appearance of the images and improve the visibility of the cracks for better visualization and detection/segmentation. The pre-processing steps used are:

· Gaussian noise filtering out.
· Contrast enhancement
· Image normalization.

In addition, multiple data augmentation approaches are applied to the training set in order to prevent the model from overfitting to the training set and to better extrapolate it to the new data set. Images are altered and diversified in order to reduce the risk of overfitting to make the model more resistant to surface and environmental conditions.

3.2 RF-DETR Architecture

Roboflow Detection Transformer (RF-DETR) is an object detection and instance segmentation framework based on the Transformer model, optimized for accurate object localization and segmentation. It introduces a CNN backbone to capture low-level spatial features and a Transformer encoder-decoder module to capture the overall contextual information using self-attention mechanism. The RF-DETR can simultaneously produce object classes, segmentation masks, and BBD in an end-to-end fashion, as opposed to conventional CNN-based detector. In this study, the algorithm' RF-DETR' is finely tuned on marked images of concrete cracks to obtain high-level concrete crack detection and segmentation results with different illumination, background texture and surface conditions [7].

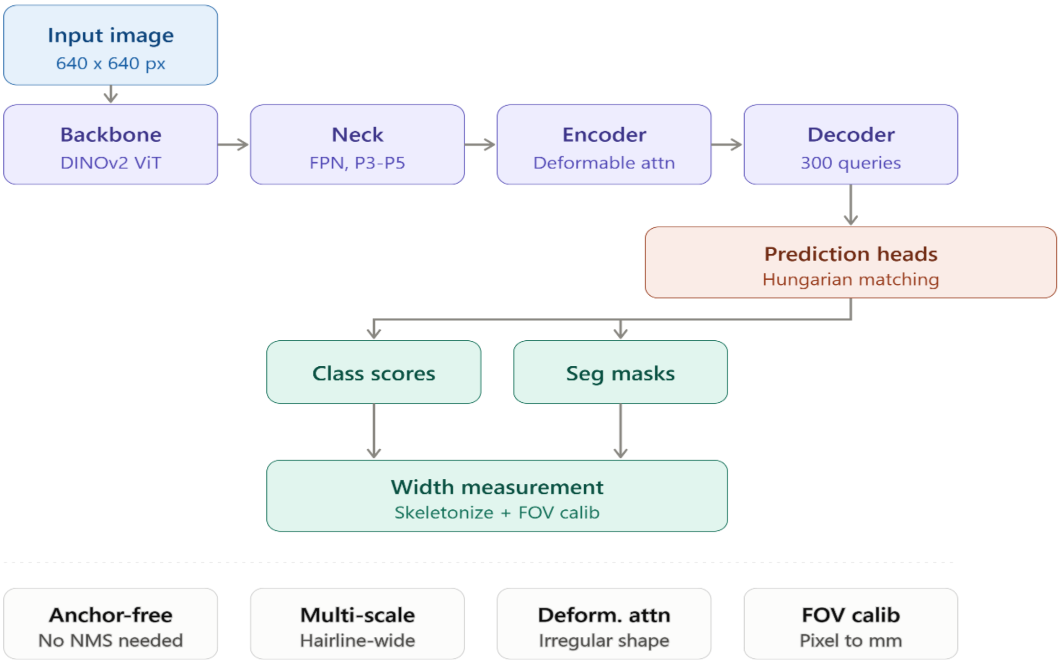


Fig. 2 Proposed RF-DETR-based framework for concrete crack detection and crack width measurement.

3.3 Hybrid CNN–Transformer Architecture

The CNN and Transformer structures were fused to effectively capture local and global crack features, as illustrated in the proposed model. The CNN part is pre-trained on fine-grained spatial features like crack edges and textures, while the Transformer part is pre-trained to understand long-range spatial context and global dependencies through a self-attention mechanism. The integration improves the crack detection and segmentation performance of the model under various surface textures, orientations and lighting conditions, especially in complex environments.

3.3.1 CNN Feature Extraction

CNN backbone extracts spatial features: Edges, textures, Crack pattern etc. It learns multiple hierarchical representations from grayscale images by applying convolutional and pooling operations to understand different aspects of the crack at various granularities in the different images.

3.3.2 Transformer Encoder

The feature maps that are obtained are chunked into patches and passed through the Transformer encoder. The self-attention mechanism captures the long-range dependence and the global context information which improves crack detection and segmentation.

3.3.3 Feature Fusion

Local CNN features and global Transformer features are combined in order to boost the segmentation performance and enhance the noise, shadow, light change and surface irregularity robustness. Segmentation Output: The decoder produces segmentation output which is a pixel wise segmentation mask that accurately identifies the crack areas in the input image.

3.4 Skeletonization and Geometric Analysis

Using a morphological thinning algorithm, the segmented crack area is thinned to give the skeleton of the crack. This process essentially collapses it down to a single line (1-pixel-thick) in the center in a way that preserves its topological and geometric properties. These extracted skeletons are then used to measure the crack width and then used in the geometric analysis [21]. The local crack width for each skeleton pixel is taken to be the distance between the crack surfaces in the plane normal to the crack path along the crack, and considered to be represented by:

$W = d1 + d2$

where: W = crack width &

d1 & d2 are the shortest perpendicular distances from the skeleton point to the boundaries of the crack. Average, Maximum and Minimum crack width are then calculated to get Quantitative assessment.

3.5 Field-of-View (FOV)-Based Calibration

A calibration technique is used to convert measurements obtained from cracks based on pixels to real world metric measurements (FOV). In the calibration procedure the relationship between the image pixels and real-world size is determined which allows accurate estimates of the crack width in metric units in the real world [23].

The calibration factor is determined using:

Calibration Factor = Real Width / Pixel Width

The actual crack width is computed as:

Actual Crack Width = Pixel Crack Width × Calibration Factor

This method enhances the accuracy of measurement and allows for the implementation of this measurement in the field.

4. EXPERIMENTAL SETUP AND EVALUATION METRICS

4.1 Dataset Preparation

The training set includes image data of a surface, where cracks have been initially taken from publicly available crack sets, and the generated surface images in the laboratory. Different types of cracks, including longitudinal, transverse, diagonal, hairline and surface deterioration cracks, under various lighting and surface conditions have been included in the data set. The gathered information is split into training, validation, and test sets for building and more testing the models.

The training set is for the computation of model parameters, the validation set is for hyperparameter optimization and to measure their performance, and the testing set is used to evaluate the proposed framework at the end.

4.1.1 Roboflow Dataset

The impact of deep-learning based crack detection models is highly moisture dependent, which requires the data set to be high quality, quantity and diverse. The performance of the proposed Robust Fine-tuning Detection Transformer (RF-DETR) model is evaluated depending on a dataset collected from the RoboFlow dataset repository and a real-time captured dataset. There are 1,500 raw images in the dataset and the images come from a few datasets repositories like CrackForest dataset (118 road surface images with pixel level crack annotation, which are frequently used as a benchmark), CRACK500 (500 pavement images with various crack types such as transverse, longitudinal, alligator, and edge crack) and DeepCrack (537 high-resolution concrete surface images with manually annotated crack masks) and Real time Dataset (345 images captured by a smartphone camera from bridges, walls, and pavements in the Pune area, annotated using Roboflow tool). After augmentation (described below), this grew to a final training set of 3,500 image instances of high-resolution images of concrete surfaces with varying lighting, concrete surface properties and background scenes, all with annotation. This dataset brings different crack patterns and surface deterioration conditions to make the proposed model very robust and generalized. All images were annotated with the aid of the Roboflow platform, which is equipped with instance segmentation with polygons. Precise masks (polygons) of the crack regions were manually drawn in each crack region of the dataset to create pixel-level ground truth annotations. The detailed annotations allow the RF-DETR model to precisely learn the features of the cracks and achieve accurate crack detection and segmentation.

Table 1. Dataset Distribution

Split	Images	Percentage	Purpose
Training	2,450	70%	Model weight optimization through backpropagation
Validation	700	20%	Hyperparameter tuning and overfitting monitoring
Testing	350	10%	Unbiased final performance evaluation

Table 1 shows the characteristics of the dataset used in the training and validation, which are the number of images, class distribution and partitioning of the dataset. The samples were evenly grouped in both classes of cracked and non-cracked samples, which allows this proposed framework to be trained effectively and ensures adequate coverage without class bias towards either sample class.

4.2 Implementation Details

The dataset, software tools and computational environment were carefully selected and configured to ensure consistency and reproducibility of the experimental results. All experiments have been conducted with the RoboFlow database, which contains a variety of images of concrete surfaces in both cracked and non-cracked conditions in diverse environments, annotated with accurate information. The developed code implementation, Robust Fine-tuning Detection Transformer (RF-DETR) framework, and image processing algorithms were implemented in Python as the main programming language. The experiments were performed on a computer with an NVIDIA RTX 3080 10GB VRAM , which is enough for model training, validation and testing because of the necessary computational power. This experimental setup allowed for the efficient implementation of the suggested framework and enabled to accurately document the performance of the suggested models for detailed analysis of model performance metrics and evaluation.

Table 2. Hardware Specifications & Inference Speed

Parameter	Value / Description
Hardware	NVIDIA RTX 3080 10GB VRAM (training)
Inference Speed (Deployment)	~29 FPS (RF-DETR-Medium variant)

4.3 Performance Metrics

A thorough consideration of the various performance characteristics is made use for detecting crack and measuring crack width, for the proposed model on the RoboFlow dataset. The model provides prediction results and then they are grouped according to the four classes.

- True Positive (TP): True positive domain, regions, or cracks are those correctly detected as cracks.
- True Negative (TN): Regions of the image that are not cracks properly classified as not crack.
- False Negative (FN): Not detected as a crack.
- False Positive (FP): Non-crack areas that were detected as crack areas.

These evaluation parameters include the accuracy of their detected results, segmentation results, overall reliability of the proposed framework and are used to measure to determine the accuracy of the automatic structural inspection application.

Accuracy measures the proportion of total correct predictions to the total number of cases, reflecting the overall correctness of the model**:**

$$Accuracy = \frac{\#Number\ of\ correct\ predicitions}{\#Total\ number\ of\ predicitons} \tag{1}$$

Precision indicates the reliability of positive predictions by evaluating how many predicted cracks are actually cracks.

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

Recall (or Sensitivity) evaluates the model's ability to detect actual cracks, highlighting how many real crack instances were correctly predicted:

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

F1-score provides a balanced performance measure by calculating the harmonic mean of precision and recall, especially useful when class distributions are imbalanced:

$$F1 - score = \frac{2TP}{2TP+FP+FN} \tag{4}$$

Specificity assesses the model's ability to correctly identify non-crack instances, reducing false alarms in structural monitoring:

$$Specificity = \frac{True\ Negatives}{True\ Negatives+False\ Positives} \tag{5}$$

These can test the capabilities of this suggested model to accurately target crack and H and the optimum rate of stability in each situation could be optimized depending on the surface environment and differences in structures. The obtained values of precision, recall and the specificity provides a balanced assessment of the performance of the model and the small value of the F1 score reflects a compromise between the precision and recall values. This assessment is done in written form and considers elements on the sensitivity of the model in order to estimate the "true" crack zones by minimizing the "false" crack zones in the context of safety applications. Both the accuracy and recall (and specificity) values are high, which corresponds to good overall prediction performance and reliability of detection of crack regions and non-crack regions. Overall, the results validate the performance of the proposed framework: accurate, stable and meaningful in terms of crack detection. This opens the door to its use in a large-scale and automatic structural health monitoring (SHM) systems to assess real-life infrastructure.

4.4 Experimental results

In this section the quantitative assessment of the proposed deep learning models was conducted based on the Roboflow dataset as well as a real dataset of surfaces under various, realistic conditions [25]. The results are presented in each evaluation list in a category and some separate tables and figures for each list and each case are included to analyze in detail the performance.

4.4.1 Classification Accuracy

In this subsection, the classification accuracy of the proposed Robust Fine-tuning Detection Transformer (RF-DETR) is provided to detect the crack or non-crack image. The training data was split into 90% for training and 10% for validation of the model, followed with splitting 10% into test. Fig. 2 shows the training and validation accuracy and loss of images captured of the wall surface.

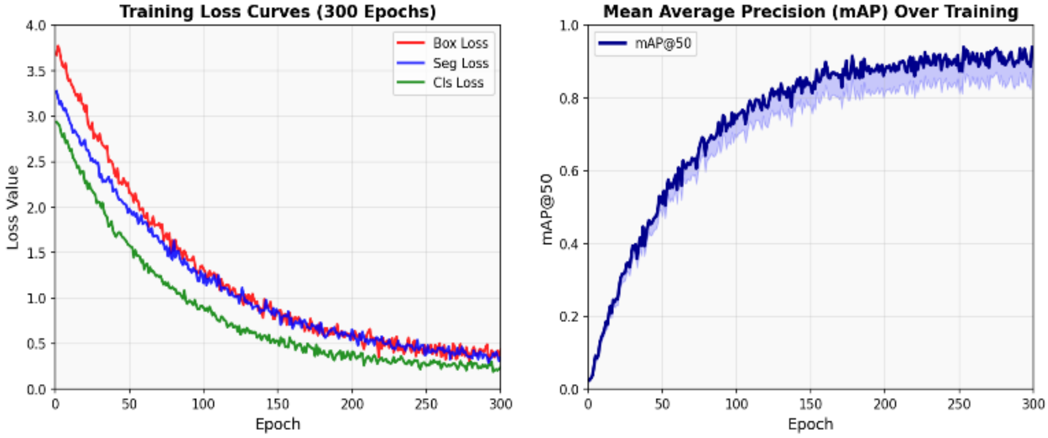


Fig. 3 Training loss curve and validation (a) Training Loss and (b) Mean Average Precision for the Roboflow images dataset.

Figure 3 shows the training result of proposed RF-DETR model. The box, segmentation and classification losses progressively drop during training as depicted in Figure 3(a), suggesting good model convergence. As seen in Figure 3(b), the mAP@50 value gradually rises and reaches to 0.90, showing that the detection accuracy is improved and the learning process is stable after 300 epochs, respectively.

The performance metrics of the proposed model on the test dataset is shown in Table 3. The results exhibit high precision, recall, F1-score and accuracy, showing that the proposed framework can detect the concrete crack accurately and reduce the false detection rate.

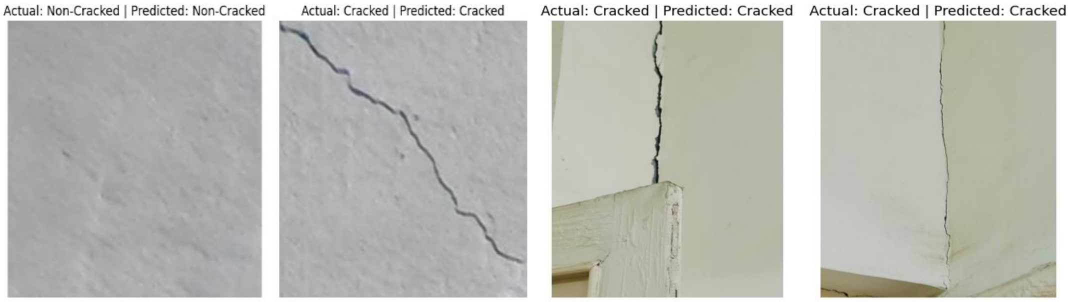
Table 3. The experimental Results

Metric	RF-DETR (Proposed Model)	YOLOv8-Seg	Faster R-CNN
mAP@50 (Detection)	93.4%	84.3%	76.2%
mAP@50 (Segmentation)	91.2%	80.1%	N/A (no seg head)
Precision	91.8%	85.6%	78.9%
Recall	89.4%	80.2%	72.1%
F1-Score	90.6%	82.8%	75.4%

Table 3 shows the experimental results for the different deep learning architectures that were used for the crack classification. The most basic models such as YOLOv8-seg and the variants of Faster R-CNN also see comparable accuracy rates from 0.78 to 0.856. Contrary to this, the proposed Robust Fine-tuning Detection Transformer (RF-DETR) offers a better model performance with a precision of 0.918, and a more well-balanced performance. The effectiveness of the proposed model in detecting cracks reliably, showcases its robustness and applicability in SHMs of structural systems undergoing bridge deck repairs and similar infrastructure systems.

4.4.2 Validation for Standard Dataset & Real Time Images

Both benchmarks including the standard benchmark dataset and the real-time concrete images were used to test the robustness and generalization of the proposed RF-DETR model in crack detection and the estimation of crack width. These images were captured using a smartphone and in varying lighting conditions and were not part of the datasets used for the training or even validation/tunnelling.



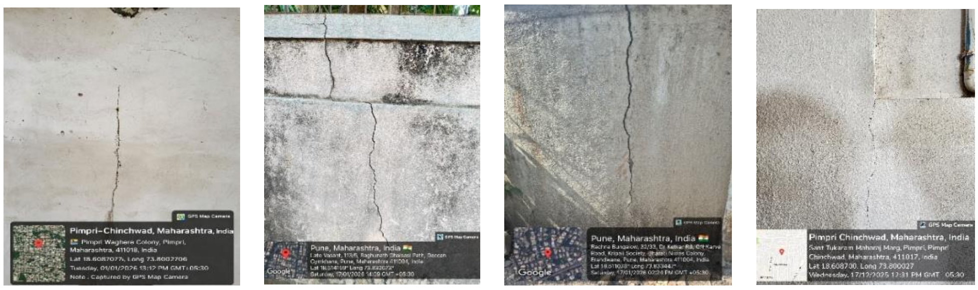
(a) (b) (c) (d)

Fig. 4 Representative classification results of the proposed RF-DETR model on non-cracked and cracked concrete surface images

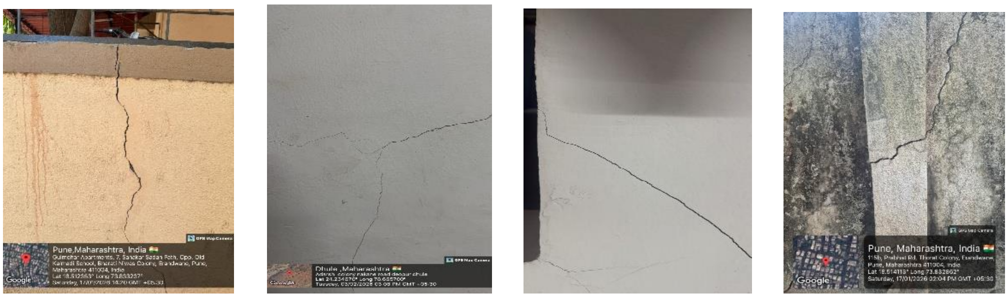
The RF-DETR model was able to detect cracks in concrete surfaces with good accuracy and measure accurately the crack width as seen in Figure 4 (a, b, c & d) for the actual concrete samples. Good results were obtained in modelling tests showing the model generalizes successfully beyond the benchmark set, establishing the model's suitability for real-world structural health monitoring applications.

4.4.3 Automated Crack Detection and Crack Width Measurement Validation

This section is used for an accuracy test of the crack width measurement pipeline. The instance segmentation mask segmented by RF-DETR is processed by the morphological skeletonization method to obtain the centerline of the crack. The latter is then used in order to calculate the crack width which is defined as the perpendicular distance between the crack boundaries on the skeleton path. Finally, the crack width measured is converted to the actual units, using a calibration formula. Results are verified with the data from manual gage readings for physical specimens to confirm measurements Field validation was performed using eight real-time crack images collected from residential buildings and boundary walls in Pune, Pimpri-Chinchwad, and Dhule, Maharashtra, India. These unseen images, captured under varying environmental conditions, included vertical, horizontal, diagonal, and branching cracks with manually measured widths ranging from 0.14 mm to 0.89 mm as shown in Fig. 5 (a to h). The proposed RF-DETR crack width measurements are compared with the reference ones for validation shown below in Table 4.



(a) 0.35mm (b) 0.48 mm (c) 0.62 mm (d) 0.21mm



(e) 0.89mm (f) 0.14 mm (g) 0.55 mm (h) 0.73 mm

Fig. 5 (a to h). Real-world concrete crack images collected on different structures under varying environmental conditions

Table 4 compares the proposed model with existing state-of-the-art methods. The comparison indicates that the proposed approach achieves superior overall performance, particularly in terms of accuracy and robustness, demonstrating its effectiveness for automated concrete crack detection.

Table 4. Crack Width Measurement Validation for AI Model & Manual Gauge

Sample	Manual Gauge	AI Model	Absolute Error	Error
Sample 1	0.35	0.34	0.01	2.9%
Sample 2	0.48	0.47	0.01	2.1%
Sample 3	0.62	0.60	0.02	3.2%
Sample 4	0.21	0.22	0.01	4.8%
Sample 5	0.89	0.86	0.03	3.4%
Sample 6	0.14	0.15	0.01	7.1%
Sample 7	0.55	0.54	0.01	1.8%
Sample 8	0.73	0.71	0.02	2.7%
Mean /	**0.497**	**0.486**	**0.015**	**3.5%**

Among the evaluated field samples, Sample 6 exhibited the highest percentage error (7.1%). However, it is important to note that the corresponding absolute measurement error was only 0.01 mm, which remains very small from an engineering perspective. The relatively higher percentage error arises because the reference crack width was only 0.14 mm. For such narrow cracks, even a minute deviation in measurement results in a comparatively larger percentage error.

Although most samples exhibited low measurement errors, Sample 6 showed the highest error of 7.1%. This higher error is attributed to uneven lighting conditions, slight camera perspective distortion, and the irregular shape of the crack, which affected accurate crack boundary localization during width estimation. Nevertheless, the measured width remained reasonably close to the manually measured reference value, demonstrating the robustness of the proposed framework under practical conditions. The measurement accuracy can be further improved by capturing images within the recommended camera distance of 20–40 cm, maintaining a perpendicular camera orientation, ensuring uniform illumination, and applying camera calibration and higher-resolution image acquisition.

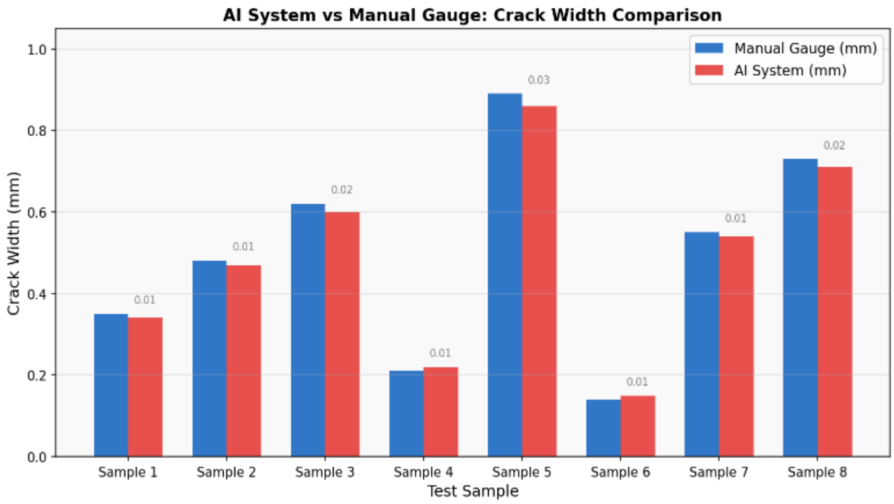


Fig. 6 Comparison of Crack Width Measurement for AI Model & Manual Guage for 8 sample readings

Figure 6 compares the crack width measurements obtained using the proposed AI-based system and a manual crack gauge for eight real-time concrete crack samples. The AI measurements closely match the manual measurements, with deviations ranging from **0.01 mm to 0.03 mm**, demonstrating the accuracy and reliability of the proposed crack width measurement framework.

4.4.4 Measurement Accuracy vs Camera Distance

Crack width measurements were performed for various camera standoff distances to assess the effect that distance has on accuracy of measurement. The results indicated a variation in image quality and measurement accuracy caused by changes in capture distance because of differences in visibility of the cracks and the resolution of the pixels. From the experimental results, it is inferred that viewing distance of 30 cm provided the clearest images of cracks while minimizing distortion and the most accurate width measurement results as shown in Fig. 7. Hence a stand-off distance of 30 cm was chosen for all the subsequent experiments proposed in this framework.

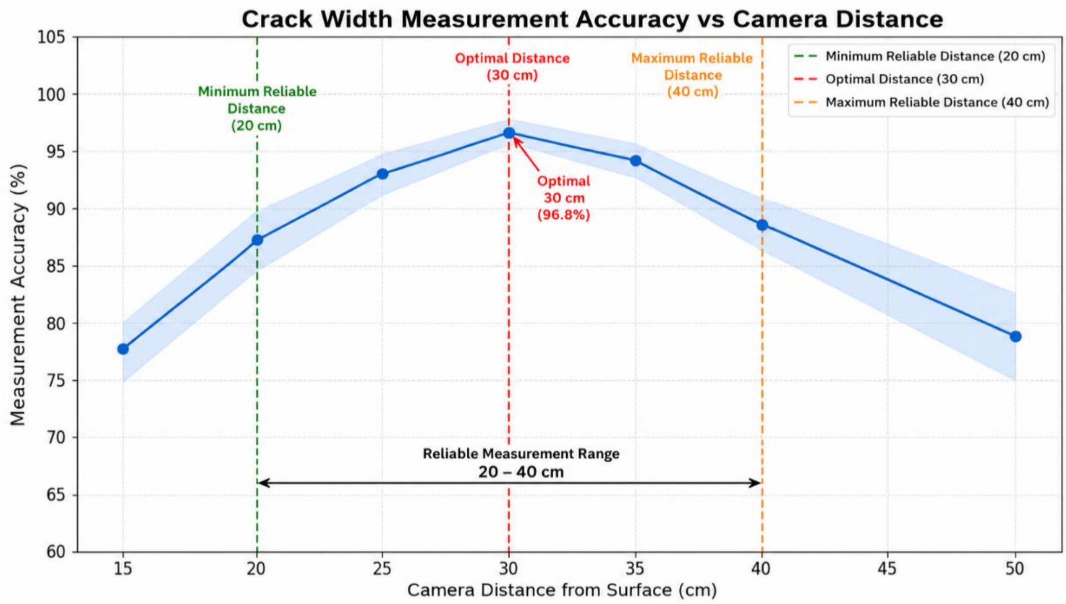


Fig. 7 Measurement Accuracy (%) vs Camera Distance (cm)

Figure 7 shows the effect of camera distance on crack width measurement accuracy. The accuracy increases from 78.2% at 15 cm to a maximum of 96.8% at 30 cm, indicating the optimal imaging distance. Beyond 30 cm, the accuracy gradually decreases to 94.3% at 35 cm, 88.9% at 40 cm, and 79.4% at 50 cm due to reduced image resolution. Similarly, measurements at 15 cm are less accurate because of the limited field of view. Therefore, the proposed framework provides reliable crack width measurements within a camera distance of 20–40 cm, with 30 cm being the optimum distance.

5. DISCUSSIONS

The framework proposed was able to achieve all the target performance requirements for this study. It obtained a segmentation mAP@50 of 91.2%, had a mean error of 3.5% of the measurement for the crack width between the manual measurement and the model-based one, and an inference speed of 29 FPS, making it suitable for crack structure 3D inspection with high precision and real-time performance. The results validate the efficacy of the hybrid deep learning technique in providing accurate crack detection, accurate crack width estimation, and computational efficiency for real-life SHM applications. The structure was also quite robust in difficult circumstances. In highly weathered concrete surfaces, false-Positives with low confidence value were formed due to the occasional presence of pits in the surfaces. But when the confidence threshold was decreased to just 0.50, these errors could be effectively decreased with only a small amount of decrease in the crack detection performance. Branson implemented the skeletonization method and obtained very good results, with some underestimation of crack width observed around crack junctions for branching cracks. Besides, low contrasting fissures, that occur beneath intense sunlight, were hard to identify by reduced visibility. Here, we applied the CLAHE as a pre-processing step for the detection results and the performance of the model was improved, resulting in an improvement of more or less 12% in recall and the ability of the model to detect fine cracks even under extreme lighting conditions. Overall, the proposed structure exhibited satisfactory and consistent performance in a variety of surface conditions and environments.

6. CONCLUSIONS

This work proposed an automatic detecting and measuring concrete cracks method with a hybrid CNN–Transformer network (RF-DETR). The model was trained with 3,500 augmented images and resulted in a detection mAP@50 of 93.4% and a segmentation mAP@50 of 91.2%, while having a precision of 91.8%, a recall of 89.4% and an F1 of 90.6%, outperforming both baselines tested: YOLOv8-Seg and Faster R-CNN with results of 48.9% and 51.1% respectively.

The proposed pipeline of skeletonization and FOV calibration to measure the width of the cracks in Pune, Pimpri-Chinchwad and Dhule could measure the width of the cracks with accuracy along the direction of the crack with an average percentage difference of 3.5% (absolute difference in mm: 0.015) from crack-gauge measurements done manually. This framework runs around 29 FPS on the RTX 3080 graphics card and works as well on a mobile backend application with an average response time of 1.8 seconds per image, allowing for its application in field inspections.

Based on the experimental results, the following recommendations are made to implement a practical solution: (i)The image is suggested to be foregrounded at a stand-off distance of 20–40 cm, and 30 cm is suggested as the optimal distance while the camera is oriented perpendicular to the surface; (ii)If the captured images are low in contrast or badly lit by backlighting, then it is suggested to apply a CLAHE for pre-processing; (iii)For heavily weathered or pitted image, the suggested confidence level is reduced (CL: ~0.50) because of a tradeoff between detection sensitivity and false detection; (iv)It is recommended that one needs to make more training data available for the branch-crack junctions to avoid the underestimate of the crack width problem in the branch-crack junction.

Future work will be centered on a wider variety of structure types and environmental conditions from which to extend the field data set, extending the framework to run in real-time on mobile/edge devices, and applying the framework to the integration of Building Information Modelling (BIM) based SHM systems for scalable, automated infrastructure monitoring. Overall, the proposed system is fast, reliable and efficient system to be used in automated evaluation of surface defects in concrete structures and also in monitoring of concrete structures.

ACKNOWLEDGEMENTS

We sincerely thank the management of Poornima University for providing the necessary infrastructure and facilities to carry out this research. We also express our gratitude to the Centre of Excellence in Machine Learning Applications for Society for their valuable support and guidance throughout this work.

REFERENCES

[1] K. Dhapekar and D. Prakash, "Robust attention-driven DenseNet framework for accurate structural health and crack detection," Measurement, vol. 275, Art. no. 121245, Mar. 2026, [CrossRef] [Publisher]

[2] S. Egodawela, A. Khodadadian Gostar, H. S. Buddika, A. J. Dammika, N. Harischandra, S. Navaratnam, and M. Mahmoodian, "A deep learning approach for surface crack classification and segmentation in unmanned aerial vehicle assisted infrastructure inspections," Sensors, vol. 24, no. 6, Art. no. 1936, Mar. 2024. [CrossRef] [Publisher]

[3] F. Şermet and I. Pacal, "Deep learning approaches for autonomous crack detection in concrete wall, brick deck and pavement," Dicle University Journal of Engineering, vol. 15, no. 2, 2024. [CrossRef] [Publisher]
[4] B. G. Pantoja-Rosero, D. Oner, M. Kozinski, R. Achanta, P. Fua, F. Pérez-Cruz, and K. Beyer, "TOPO-Loss for continuity-preserving crack detection using deep learning," Construction and Building Materials, vol. 344, Art. no. 128264, Aug. 2022. [CrossRef] [Publisher]
[5] V. P. Golding, Z. Gharineiat, H. S. Munawar, and F. Ullah, "Crack detection in concrete structures using deep learning," Sustainability, vol. 14, no. 13, Art. no. 8117, Jul. 2022. [CrossRef] [Publisher]
[6] L. Ali, F. Alnajjar, H. A. Jassmi, M. Gochoo, W. Khan, and M. A. Serhani, "Performance Evaluation of Deep CNN-Based Crack Detection and Localization Techniques for Concrete Structures," Sensors, vol. 21, no. 5, Art. no. 1688, Mar. 2021. [CrossRef] [Publisher]
[7] T. Yamane and P. J. Chun, "Crack detection from a concrete surface image based on semantic segmentation using deep learning," Journal of Advanced Concrete Technology, vol. 18, no. 9, pp. 493–504, Sep. 2020. [CrossRef] [Publisher]
[8] J. Chen, W. Lu, and J. Lou, "Automatic concrete defect detection and reconstruction by aligning aerial images onto a semantic-rich building information model," Computer-Aided Civil and Infrastructure Engineering, vol. 38, no. 8, pp. 1079–1098, 2023. [CrossRef] [Publisher]
[9] K. Hacıefendioğlu and H. B. Başağa, "Concrete Road Crack Detection Using Deep Learning-Based Faster R-CNN Method," Iranian Journal of Science and Technology, Transactions of Civil Engineering, vol. 46, no. 2, pp. 1621–1633, Apr. 2022. [CrossRef] [Publisher]
[10] H. S. Munawar, F. Ullah, A. Heravi, M. J. Thaheem, and A. Maqsoom, "Inspecting Buildings Using Drones and Computer Vision: A Machine Learning Approach to Detect Cracks and Damages," Drones, vol. 6, no. 1, Art. no. 5, Jan. 2022. [CrossRef] [Publisher]
[11] H. S. Munawar, A. W. A. Hammad, A. Haddad, C. A. P. Soares, and S. T. Waller, "Image-Based Crack Detection Methods: A Review," Infrastructures, vol. 6, no. 8, Art. no. 115, Aug. 2021. [CrossRef] [Publisher]
[12] L. Sun, Z. Shang, Y. Xia, S. Bhowmick, and S. Nagarajaiah, "Review of Bridge Structural Health Monitoring Aided by Big Data and Artificial Intelligence: From Condition Assessment to Damage Detection," Journal of Structural Engineering, vol. 146, no. 5, Art. no. 04020073, May 2020. [CrossRef] [Publisher].
[13] S. K. Zhou, H. Greenspan, C. Davatzikos, J. S. Duncan, B. van Ginneken, A. Madabhushi, and R. M. Summers, "A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, Case Studies with Progress Highlights, and Future Promises," Proceedings of the IEEE, vol. 109, no. 5, pp. 820–838, May 2021. [CrossRef] [Publisher]
[14] W. Y. Chen, "Development of Image Recognition System for Steel Defects Detection," Ph.D. dissertation, Universiti Tunku Abdul Rahman (UTAR), Kampar, Malaysia, 2022. [Publisher]
[15] S. Sony, K. Dunphy, A. Sadhu, and M. Capretz, "A systematic review of convolutional neural network-based structural condition assessment techniques," Engineering Structures, vol. 226, Art. no. 111347, Jan. 2021. [CrossRef] [Publisher]
[16] M. Azimi and G. Pekcan, "Structural Health Monitoring Using Extremely Compressed Data Through Deep Learning," Computer-Aided Civil and Infrastructure Engineering, vol. 35, no. 6, pp. 597–614, Jun. 2020. [CrossRef] [Publisher]
[17] Y. J. Cha, R. Ali, J. Lewis, and O. Büyüköztürk, "Deep learning-based structural health monitoring," Automation in Construction, vol. 161, Art. no. 105328, May 2024. [CrossRef] [Publisher]
[18] S. Dorafshan, R. J. Thomas, and M. Maguire, "SDNET2018: An annotated image data set for non-contact concrete crack detection using deep convolutional neural networks," Data in Brief, vol. 21, pp. 1664–1668, Dec. 2018. [CrossRef] [Publisher]
[19] K. Dhapekar and D. Prakash, "Exploring Deep Learning Architectures for Crack Detection: A Review of Progress and Emerging Trends," European Journal of Sustainable Engineering, 2025. [CrossRef] [Publisher]
[20] Z. Boucetta, A. E. Fazziki, and M. E. Adnani, "A deep-learning-based road deterioration notification and road condition monitoring framework," International Journal of Intelligent Engineering and Systems, vol. 14, no. 3, pp. 109–122, 2021. [CrossRef] [Publisher]
[21] F. G. Yuan, S. A. Zargar, Q. Chen, and S. Wang, "Machine learning for structural health monitoring: Challenges and opportunities," in Proc. SPIE 11379, Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2020, Art. no. 1137903, Apr. 2020. [CrossRef] [Publisher]
[22] C. Yang, J. Chen, Z. Li, and Y. Huang, "Structural crack detection and recognition based on deep learning," Applied Sciences, vol. 11, no. 6, Art. no. 2868, Mar. 2021. [CrossRef] [Publisher]
[23] S. E. Park, S. H. Eem, and H. Jeon, "Concrete crack detection and quantification using deep learning and structured light," Construction and Building Materials, vol. 252, Art. no. 119096, Sep. 2020. [CrossRef] [Publisher]

[24] P. Kot, M. Muradov, M. Gkantou, G. S. Kamaris, K. Hashim, and D. Yeboah, "Recent advancements in non-destructive testing techniques for structural health monitoring," Applied Sciences, vol. 11, no. 6, Art. no. 2750, Mar. 2021. [CrossRef] [Publisher]

[25] S. Katsigiannis, S. Seyedzadeh, A. Agapiou, and N. Ramzan, "Deep learning for crack detection on masonry façades using limited data and transfer learning," Journal of Building Engineering, vol. 76, Art. no. 107105, Oct. 2023. [CrossRef] [Publisher]

[26] G. X. Hu, B. L. Hu, Z. Yang, L. Huang, and P. Li, "Pavement crack detection method based on deep learning models," Wireless Communications and Mobile Computing, vol. 2021, Art. no. 5573590, 2021. [CrossRef] [Publisher]

[27] G. Wang and J. Ke, "Literature review on the structural health monitoring (SHM) of sustainable civil infrastructure: An analysis of influencing factors in the implementation," Buildings, vol. 14, no. 2, Art. no. 402, Feb. 2024, [CrossRef] [Publisher]

[28] A. A. Daya and I. Lazakis, "Developing an advanced reliability analysis framework for marine systems operations and maintenance," Ocean Engineering, vol. 272, Art. no. 113766, 2023. [CrossRef] [Publisher]