Review of Computer Engineering Research
2026 Vol. 13, No. 4, pp. 220-240
ISSN(e): 2410-9142
ISSN(p): 2412-4281
DOI: 10.18488/76.v13i4.5371
© 2026 Conscientia Beam. All Rights Reserved.

Attention-augmented deep architectures for high-precision cement surface crack detection

Deepti Vadicherla[1+]
Poonam Gupta[2]

[1,2]*Department of Computer Science and Engineering, G H Raisoni University, Amaravati, India.*
[2]*Department of Information Technology, G H Raisoni College of Engineering and Management, Pune, India.*
[1]*Email: deepti.vadicherla@gmail.com*
[2]*Email: poonam.gupta@raisoni.net*



(+ Corresponding author)

Article History
Received: 7 April 2026
Revised: 28 July 2026
Accepted: 23 August 2026
Published: 20 November 2026

Keywords
Automated visual inspection
Cement surface crack detection
Convolutional block attention module
Deep learning
InceptionV3
Structural health monitoring.

ABSTRACT

This study aims to improve the reliability and efficiency of automated cement surface crack detection by evaluating attention-augmented deep learning architectures for structural health monitoring. Accurate crack identification is essential for maintaining infrastructure safety, reducing manual inspection errors, and supporting timely maintenance decisions. Traditional visual inspection methods are time-consuming, subjective, and prone to human fatigue, whereas conventional convolutional neural networks misclassify complicated surface textures, stains, and variations in illumination as structural cracks. To overcome these limitations, we present and compare three deep learning architectures under identical experimental conditions. Performance was achieved with a conventional CNN based on InceptionV3, with additional performance improvements obtained by incorporating a CBAM module into InceptionV3 and into VGG16. The convolutional block attention module applies spatial and channel attention in series to improve representations and direct the network to regions of interest while suppressing background noise. All models were trained and tested on a balanced set of 4000 high-resolution images of cement surfaces with both cracked and non-cracked samples. Results confirmed that the CBAM + InceptionV3 architecture provided the highest overall performance, with 99.17% accuracy and 99.33% sensitivity, and a reduction in false positives relative to the conventional CNN. In addition, the CBAM + VGG16 architecture achieved comparable accuracy with substantially lower computational requirements and was therefore well-suited for applications with limited resources. The practical implication of this study is that attention-based feature refinement can enhance decision reliability, reduce inspection uncertainty, and support the development of efficient automated systems for civil infrastructure monitoring.

Contribution/Originality: This study contributes to the existing literature by providing a controlled comparative analysis of baseline and CBAM-enhanced CNN architectures for cement surface crack detection. This study uses an attention-based feature-refinement methodology and documents that CBAM improves precision, reduces false positives, and supports reliable deployment-oriented structural health-monitoring applications.

1. INTRODUCTION

Early detection of surface cracks in concrete and cement structures is critical for achieving structural safety, extending service life, and reducing costs of civil infrastructure. Surface cracking typically represents one of the first visual indications of material deterioration, stress concentrations, and potential structural deficiencies. Failure to detect such defects at an early stage may ultimately result in progressive increases in the magnitude of damage and risk to structural safety. Conventional methods for crack inspection generally rely on manual visual examination or conventional techniques for image analysis, such as thresholding, edge detection, and morphological operations.

© 2026 Conscientia Beam. All Rights Reserved.

However, these methods are generally laborious, time-consuming, subjective, and highly sensitive to variations in lighting conditions, surface texture, noise, and observer bias [1, 2].

Deep learning with convolutional neural networks has dramatically improved automated crack detection by learning discriminative features directly from raw images, without the need for manual feature engineering [3-5]. This leads to substantially greater accuracy, robustness, and adaptability than conventional methods for many applications of visual inspection. However, the strong capabilities for feature learning remain insufficient to distinguish true structural cracks from highly complex textures, stains, shadows, and surface irregularities. As a result, performance is particularly poor for images of concrete surfaces in which cracks are typically narrow, irregular, and difficult to differentiate from non-crack conditions. Finally, classification results obtained with conventional CNN models frequently include false-positive calls or fail to identify the most informative aspects of cracks [6].

To overcome this challenge, attention mechanisms have been incorporated into deep learning models to improve feature selection and spatial focusing. In particular, our lightweight and effective convolutional block attention module (CBAM) sequentially applies channel and spatial attention to extract optimal features. Channel attention enables selection of the most informative feature maps, and spatial attention focuses on regions of interest corresponding to cracks in the image. As a result, CBAM reduces the influence of redundant background information and increases the importance of structurally relevant features associated with cracks, thereby increasing the reliability of CNN-based crack classification [7].

The novelty of this work lies in providing a controlled, deployment-oriented comparative evaluation of baseline and attention-enhanced CNNs for detecting cracks on cement surfaces. Rather than evaluating a single model or reporting results obtained under different experimental conditions, we compared CNN + InceptionV3, CBAM + InceptionV3, and CBAM + VGG16 on a common dataset, with identical preprocessing, training, and evaluation conditions. In addition, we assessed not only classification accuracy, but also sensitivity, specificity, confusion-matrix behavior, the stability of results obtained through cross-validation, model complexity, the time required for inference, and the statistical significance of the results. Consequently, our major objective was not only to achieve high detection accuracy, but also to determine whether attention mechanisms increased the reliability of decisions, reduced false-alarm rates, and facilitated application to practical structural health-monitoring systems.

The major contributions of this work are:

A controlled experimental approach is used to compare conventional and CBAM-based systems for detecting surface cracks on cement. We evaluate the effects of channel and spatial attention on improving crack features and reducing false-positive results. Model performance is quantified with 5-fold cross-validation and statistical tests for significance. Finally, we quantify factors limiting implementation, such as inference time, model size, number of parameters, and training efficiency. This allows evaluation of the feasibility of applying the approach to real inspection systems.

The remainder of this work is organized as follows. Section 2 provides an overview of existing work on conventional, CNN-based, and attention-enhanced methods for crack detection. Section 3 details our proposed methodology with respect to the baseline CNN architecture, integration of CBAM modules, training parameters, and implementation environment. Details of the dataset, preprocessing, augmentation strategies, experimental conditions, and performance evaluation are provided in Section 4. Experimental results are given in Section 5 for comparisons of performance, analysis of confusion matrices, results of cross-validation, statistical validation of performance, measures of model complexity, and results of an ablation study. Section 6 discusses the significance, limitations, and relationship to recent work, and the final section presents our overall conclusions and outlines directions for future research.

© 2026 Conscientia Beam. All Rights Reserved.

2. RELATED WORK

Automated crack detection has attracted considerable interest in structural health monitoring because surface cracks reflect incipient degradation of concrete and cement-based infrastructure. Earlier methods primarily employed traditional image-processing techniques for crack detection, including edge detection, thresholding, morphological operations, and manually developed textural features. Although these methods are computationally efficient and relatively easy to implement, performance was strongly dependent on image quality, illumination conditions, background texture, and the choice of threshold. Consequently, these methods generally failed for thin, discontinuous, poorly illuminated, or morphologically similar cracks, stains, joints, shadows, or textures of rough surfaces [2, 3]. This experience motivated the development of data-driven approaches capable of automatically learning discriminative features for cracks from images.

Deep learning and, in particular, convolutional neural networks have substantially improved the performance of automated crack detection and eliminated the need for hand-crafted features [4, 5]. These models directly learned hierarchical representations of spatial information from raw images and achieved substantially greater accuracy for classification and for delineation of crack locations compared to conventional methods. Results of application to controlled data sets were highly accurate, but most methods provided little information concerning rates of false-positive results, stability of generalization, costs of computation or difficulties of implementation. In addition, training of conventional convolutional neural networks on benchmark data resulted in substantial learning of global patterns of appearance, but generally did not provide maximum sensitivity for regions of the image most relevant to detection of cracks [6, 8]. This ultimately contributed to errors of misclassification in the presence of noise, stains, textures of irregular surfaces or appearances similar to those of cracks.

Backbone architectures such as InceptionV3 and VGG16 have become widely used in computer vision applications because of their strong capability for feature extraction. InceptionV3 is well-suited to capture multi-scale visual information and thus is effective for detection of cracks of varying widths, orientations, and shapes. In contrast, VGG16 provides a simpler sequential architecture that is more readily interpretable and applicable to certain practical conditions [9, 10]. However, these network architectures were originally developed for general purposes of image classification and not for detailed analysis of structural defects. Consequently, direct application to images of cement surfaces resulted in excessive extraction of redundant information representing background features in addition to features of cracks. This is of particular importance for detection of cracks, in which the area of defect occupies only a small fraction of the image and the representation is dominated largely by details of the surrounding texture.

Recent efforts to improve CNN-based crack detection have included multi-scale learning, segmentation networks, lightweight models, and transformer-based methods. For example, multi-scale CNNs enabled the detection of cracks at different spatial scales, whereas segmentation networks provided pixel-level localization of cracks [11, 12]. YOLO-based and lightweight CNN models also facilitated real-time or edge implementations of crack inspection systems [13-15]. However, these approaches generally required more complex annotations, larger training datasets, and/or additional computing resources. In addition, some lightweight models reduced the richness of features, whereas segmentation-based methods required extensive and time-consuming preparation of pixel-level ground-truth images.

Attention mechanisms have become valuable for increasing the reliability of CNN-based crack detection. They enable a network to emphasize informative features and suppress less relevant responses to background noise. The convolutional block attention module of Woo et al. [7] is especially well-suited for this task, as it incorporates both channel and spatial attention. Channel attention identifies which feature maps are most important for discrimination of cracks, whereas spatial attention highlights locations of crack-relevant information within the image. This degree of refinement is particularly helpful for detection of cracks that are generally narrow, irregular, and superimposed on backgrounds of high noise or complex texture.

© 2026 Conscientia Beam. All Rights Reserved.

Recent results confirm that attention-based improvements in selectivity and reductions in sensitivity to background noise significantly increase the performance of crack detection [13, 16-20]. However, most of these studies evaluated a single architectural design, used different experimental conditions, or failed to provide a controlled comparison between baseline and attention-based models. Consequently, it was difficult to determine whether performance improvements resulted from the attention mechanism per se, the architecture of the network, the method of dividing training and test data, or the mode of training. In addition, performance was reported in terms of accuracy only, whereas practical systems for inspection of cracks require a more complete assessment of precision, recall, specificity, patterns of false-positive responses, duration of inference, size of the model, stability of results obtained with cross-validation, and results of experiments with ablation of attention modules. Table 1 presents a critical comparison of recent crack detection studies and identifies research gaps.

Table 1. Critical comparison of recent crack detection studies and identified research gaps.

Ref.	Study Focus	Method Used	Key Strength	Limitation / Research Gap
Golding, et al. [4] and Yamane and Chun [5]	Review of CNN-based crack detection	Deep learning-based crack detection review	Shows CNNs outperform traditional image-processing methods	Mostly summarizes existing methods; limited focus on attention-based reliability and deployment metrics
Yuan, et al. [6] and Ozdemir, et al. [8]	Attention-enhanced CNN models	Modified attention-based deep networks / ConvNeXt-based models	Improves feature representation and crack discrimination	Limited controlled comparison across different CNN backbones
Woo, et al. [7]	Attention mechanism	CBAM: channel and spatial attention	Provides lightweight feature refinement using channel and spatial attention	Originally proposed for general vision tasks; requires domain-specific evaluation for cement crack detection
Dong, et al. [13] and Yadav, et al. [21]	Detection under complex backgrounds	YOLOv8 / CNN-transformer hybrid models	Better robustness in complex surface conditions	Higher model complexity; deployment cost and false-positive behavior not always deeply discussed
Xiang, et al. [11] and Dais, et al. [12]	Crack segmentation	Super-resolution and segmentation networks	Provides pixel-level crack localization	Requires pixel-level annotations, which are costly and time-consuming
Munawar, et al. [14]; Raza, et al. [15]; Yu, et al. [22] and Isong [23]	Lightweight crack detection	Lightweight CNN / edge-based models	Suitable for real-time or low-resource deployment	May reduce feature richness and crack discrimination capability
Present study	Attention-augmented cement crack classification	CNN + InceptionV3, CBAM + InceptionV3, CBAM + VGG16	Controlled comparison with accuracy, precision, specificity, cross-validation, model complexity, inference time, and ablation study	Focused on binary classification; severity estimation and segmentation are future extensions

A critical review of the existing literature reveals that three major limitations remain. First, conventional image-processing methods are sensitive to illumination and surface noise. Second, although conventional CNN architectures achieve high accuracy, they fail to explicitly direct the network to learn discriminative features relevant to cracks and thus misclassify crack-like patterns as background noise. Third, recent approaches based on attention mechanisms or hybrid architectures generally lack controlled comparisons between different backbone architectures and are inadequately evaluated with respect to potential applications. Consequently, the inclusion of CBAM was required for

© 2026 Conscientia Beam. All Rights Reserved.

this work because it implements a computationally efficient strategy for increasing discrimination at both the level of individual channels and that of spatial relationships.

In light of these limitations, the present work compares three architectures under identical experimental conditions: CNN + InceptionV3, CBAM + InceptionV3, and CBAM + VGG16. This controlled comparison provides insight into the true utility of attention-based feature refinement for the detection of cracks on concrete surfaces. In contrast to previous reports that emphasize only high accuracy, our evaluation included analysis of classification performance, behavior of the confusion matrix, stability of cross-validation results, complexity of the models, speed of inference, efficiency of training, magnitude of statistical significance, and results of an ablation study. Consequently, our review confirms the need for an attention-based, deployment-conscious approach to concrete crack detection.

3. DATA AND METHODOLOGY

We applied a supervised deep learning-based image classification strategy for the automated detection of cracks on cement surfaces. Our primary method employed convolutional neural networks with integrated convolutional block attention modules. This enabled automatic extraction of hierarchical spatial representations from images of surfaces and selective enhancement of the most relevant channel- and spatial-based crack features. In contrast to most prior studies, in which a single architecture was evaluated or results obtained with different experimental conditions were reported, we compared three architectures under identical conditions of data preprocessing, training, and evaluation. This level of control enabled us to attribute most differences in performance to architectural modifications rather than to variations in data-processing or training methods.

The present approach differs from previous studies on CNN-based crack detection in three major respects. First, it compares baseline and attention-enhanced methods under identical experimental conditions to facilitate fair comparisons. Second, it tests the effectiveness of CBAM in minimizing false positives through enhanced attention to both channel and spatial features. Third, it provides deployment-related measures of inference time, model size, number of parameters, training efficiency, cross-validation stability, and the extent of ablation effects, rather than relying solely on measures of accuracy.

3.1. Dataset Description and Source

The dataset used in this study consists of cement surface images categorized into two classes: cracked and non-cracked. The images were obtained from the publicly available Surface Crack Detection dataset available on Kaggle: https://www.kaggle.com/datasets/arunrk7/surface-crack-detection. The dataset was further filtered and organized for the present study. A total of 4,000 images were used, consisting of 2,000 cracked images and 2,000 non-cracked images, thereby maintaining class balance. All images were converted to RGB format and resized to $224 \times 224 \times 3$ pixels before being supplied to the models. The dataset includes cement surface images with different texture patterns, lighting variations, crack directions, and background irregularities, which helps evaluate model robustness.

3.2. Data Preprocessing and Augmentation

All input images were resized to a fixed dimension of $224 \times 224 \times 3$ pixels to conform to the requirements for input to InceptionV3 and VGG16. Image pixel values were normalized to increase numerical stability and accelerate model convergence. To minimize overfitting and maximize generalization, we applied real-time data augmentation to training data only. Methods of data augmentation included horizontal flipping, random rotations, adjustments of brightness, zoom factors, and alterations of width and height. Data augmentation was not applied to validation and test data to ensure an unbiased assessment of model performance.

The resulting dataset was partitioned into training and test subsets with a stratified allocation to maintain equal representation of images with and without cracks. Training included 2,800 images, and testing included 1,200 images. Use of a portion of the training data for validation during model training allowed continuous monitoring of the rate

© 2026 Conscientia Beam. All Rights Reserved.

of convergence and provided a basis for early termination. Finally, five-fold cross-validation provided additional confirmation of the reliability of model generalization and further reduced concerns related to potential bias of data-split methods.

3.3. Overall Analytical Framework

The proposed crack-detection framework follows an end-to-end supervised-learning pipeline. Let the input image be represented as:

$$X \in R^{H \times W \times C} \quad (1)$$

Where H, W, and C denote image height, width, and number of channels, respectively. In this study, each input image is represented as follows:

$$X \in R^{224 \times 224 \times 3} \quad (2)$$

The model learns a mapping function.

$$f_\theta(X) \rightarrow y \quad (3)$$

where f_θ represents the deep learning model with trainable parameters θ, and y ∈ {0,1} denotes the binary class label, where 0 represents non-cracked images, and 1 represents cracked images.

The overall workflow involves image preprocessing, feature extraction with CNN backbones, optional refinement of features with attention-based mechanisms, and final binary classification with dense layers and sigmoid activation. The same workflow was applied to all three models for an unbiased comparison. The overall design of the proposed cement surface crack detection system is illustrated in Figure 1 and includes image preprocessing, feature extraction with CNNs, optional application of CBAM-based attention mechanisms, binary classification, and assessment of performance.

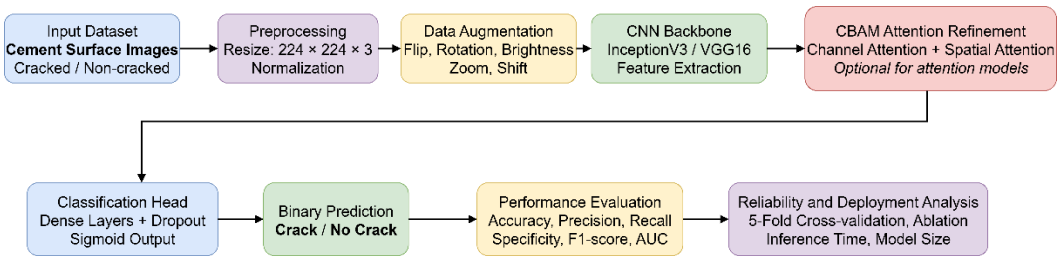


Figure 1. Overall framework of the proposed cement surface crack detection system.

3.4. Baseline CNN + InceptionV3 Model

The first model is a baseline convolutional architecture with an InceptionV3 backbone. This network was selected to exploit parallel operations that produce multiscale representations of spatial information, which are well suited to the detection of cracks with varying widths, orientations, and surface conditions. The network was trained with initial weights derived from a model pretrained on ImageNet, thus providing the benefits of transfer learning. Initial training with frozen lower layers preserved general representations of visual information, whereas fine-tuning of upper layers resulted in learning crack-specific features.

After feature extraction, the resulting feature maps were processed through fully connected dense layers with ReLU activation. Dropout regularization reduced overfitting and was followed by a sigmoid output layer for binary classification. This represents our non-attention baseline for comparison with models incorporating CBAM-based feature refinement. The architecture of this baseline CNN + InceptionV3 model is shown in Figure 2. It thus represents our non-attention reference for quantifying the effect of CBAM-based feature refinement.

© 2026 Conscientia Beam. All Rights Reserved.

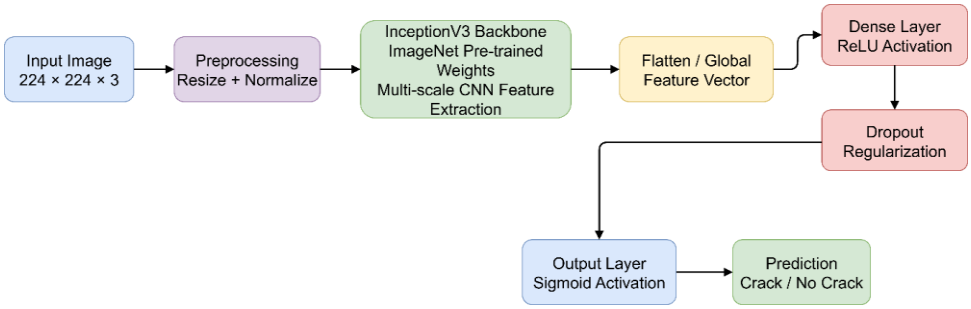


Figure 2. Architecture of the baseline CNN + InceptionV3 model for cement surface crack classification.

3.5. CBAM + InceptionV3 Model

The second model integrates CBAM with the InceptionV3 backbone. The purpose of adding CBAM is to improve feature discrimination by helping the network focus on crack-relevant information while suppressing irrelevant background variations. CBAM performs two sequential operations: channel attention and spatial attention.

The channel attention mechanism can be represented as follows:

$$F' = M_c(F) \otimes F \qquad (4)$$

where F is the input feature map, $M_c(F)$ is the channel attention map, and $\otimes$ denotes element-wise multiplication. This operation highlights important feature channels.

The spatial attention mechanism is then applied as:

$$F'' = M_s(F') \otimes F' \qquad (5)$$

where $M_s(F')$ is the spatial attention map and F'' is the refined feature representation. This operation helps the model focus on crack-specific regions in the image.

The refined feature maps are passed to the same classification head used in the baseline model. This allows the study to isolate the effect of CBAM and determine whether attention-based refinement improves prediction reliability and reduces false positives.

The architecture of the proposed CBAM + InceptionV3 model is shown in Figure 3. CBAM is inserted after the final convolutional block of InceptionV3 to refine channel-level and spatial-level features before classification.

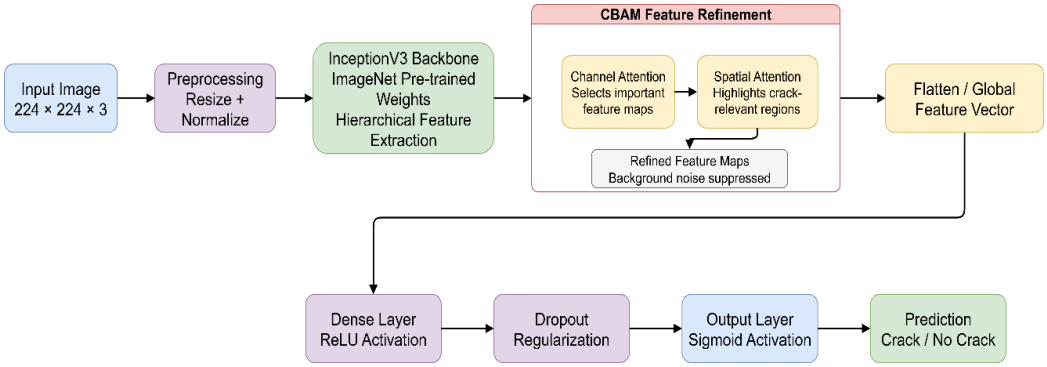


Figure 3. Proposed CBAM-integrated InceptionV3 architecture for cement surface crack detection.

3.6. CBAM + VGG16 Model

The third architecture incorporates CBAM with a simpler sequential design of repeated 3×3 convolutional layers and max-pooling. While less complex than many modern designs, it remains well suited for testing whether attention mechanisms improve feature representation with a relatively simple CNN architecture. Including CBAM after the last convolutional block allowed attention refinement of high-level semantic representations before classification.

This model was included to evaluate the balance among accuracy, complexity, inference speed, and suitability for deployment. In particular, it demonstrated the potential for achieving competitive performance with a relatively

© 2026 Conscientia Beam. All Rights Reserved.

lightweight, attention-enhanced architecture for structural inspection applications under conditions of limited resources.

3.7. Training Objective and Loss Function

The models were trained as binary classifiers using the binary cross-entropy loss function. For a given input image X_i with true label y_i, the predicted probability is denoted as:

$$\hat{y}_i = f_\theta(X_i) \quad (6)$$

The binary cross-entropy loss is defined as:

$$L = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \quad (7)$$

where N is the total number of training samples, y_i is the true class label, and $\hat{y}_i$ is the predicted probability of the crack class.

The training objective is to minimize the loss function by optimizing model parameters θ.

$$\theta^* = arg\min_\theta L(y, \hat{y}) \quad (8)$$

The Adam optimizer was used because of its adaptive learning capability and stable convergence during deep neural network training.

3.8. Training Configuration and Hyperparameters

All models were implemented using TensorFlow and Keras. The same training settings were used for all architectures to maintain fairness in comparison. Table 2 presents the training configuration and hyperparameter settings used for all evaluated models.

Table 2. Training configuration and hyperparameter settings.

Parameter	Value
Optimizer	Adam
Learning rate	0.0001
Loss function	Binary cross-entropy
Batch size	32
Maximum epochs	10
Input image size	$224 \times 224 \times 3$
Activation in output layer	Sigmoid
Regularization	Dropout
Validation strategy	Validation split from training data
Generalization check	Five-fold cross-validation

Model checkpointing and early stopping were used to preserve the best-performing model and reduce overfitting. Although the maximum training duration was set to 10 epochs, the models generally converged within the first five to seven epochs due to transfer learning and stable attention-guided feature refinement.

3.9. Bias and Limitations of the Dataset

The dataset is well balanced with respect to class distribution, but several limitations should be acknowledged. First, binary labels were used, and no discrimination was made among crack widths, severities, depths, or types. Second, because labels were assigned at the level of the image and not at the level of individual pixels, quantitative assessment of precise crack localization was precluded. Third, some images included surface stains, shadows, coarse textures, or non-structural features that mimicked cracks and potentially resulted in classification ambiguity. Finally, because the dataset is publicly available and widely used for studies of crack detection, very high performance may

© 2026 Conscientia Beam. All Rights Reserved.

represent an extent of benchmark saturation. As a result, we report results of cross-validation, ablation analysis, and evaluation of false-positive classifications to provide additional support for the reliability of our findings.

3.10. Implementation Environment

All experiments were conducted on a workstation equipped with an Intel Core i7 12th-generation processor, an NVIDIA RTX 3080 GPU with 12 GB of VRAM, and 32 GB of RAM. The software environment included Windows 11 Pro, Python 3.10, TensorFlow 2.12, Keras 2.12, OpenCV 4.7, Matplotlib, and Seaborn. The same computational environment was used for all models to ensure consistency in training time, inference time, and deployment-related comparisons.

4. EXPERIMENTAL SETUP AND RESULT ANALYSIS

The present work describes the experimental design, data preparation, methods of evaluation, and results of performance analysis for our proposed models for the detection of surface cracks in cement. The primary objective of the experiments was to determine whether attention-based refinement of features increases the reliability of classification compared to a conventional CNN architecture. We implemented and compared three models under identical experimental conditions: CNN + InceptionV3, CBAM + InceptionV3, and CBAM + VGG16. Our evaluation extended beyond measures of classification accuracy to include estimates of sensitivity and specificity, behavior with respect to false positives, model complexity, time required for inference, stability of results obtained with cross-validation, levels of statistical significance, and contributions to overall performance attributable to the attention modules.

4.1. Dataset

The dataset used in this study consists of cement surface images categorized into two classes: cracked and non-cracked. Dataset details are presented in Section 3. Figure 4 presents representative samples from the dataset.

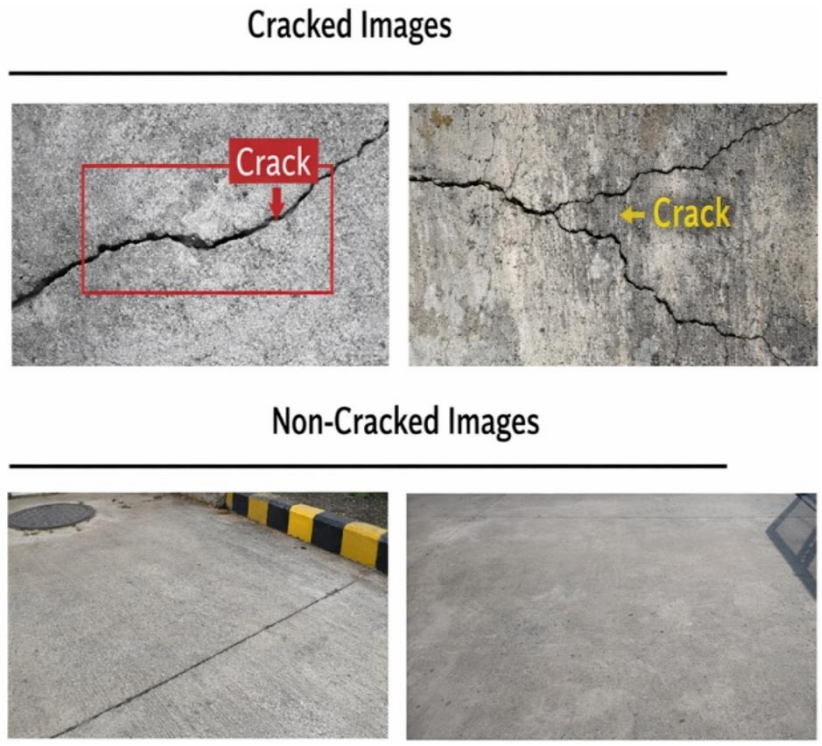


Figure 4. Representative samples of cracked and non-cracked cement surface images used in the study.

© 2026 Conscientia Beam. All Rights Reserved.

As shown in Figure 4, the dataset contains visually diverse samples, which makes it suitable for evaluating the models' ability to distinguish true crack patterns from background surface variations.

4.2. Data Preprocessing and Augmentation

All images were resized to $224 \times 224 \times 3$ pixels to conform to the input requirements of InceptionV3 and VGG16. Pixel values were normalized to increase numerical stability during training. To minimize overfitting and maximize generalization, real-time augmentation was performed only on training data. This included horizontal flipping, random rotation of ±15 degrees, variations in brightness of ±20%, zoom factors of ±10%, and shifts in width and height of ±5%. No augmentation was applied to images used for validation or testing, and reported results represent performance on unaltered data. The dataset was split using a stratified 80:20 training-to-test partition. The resulting training data included 2,800 images: 1,400 images of cracked and 1,400 images of non-cracked conditions. The test data included 1,200 images: 600 images of cracked and 600 images of non-cracked conditions. Validation data were selected from the training set to monitor convergence, implement early stopping, and record model checkpoints. Finally, five-fold cross-validation was used to assess generalization stability and minimize the likelihood that observed performance resulted from an unusually favourable single test/train split.

4.3. Computational Environment and Evaluation Metrics

All experiments were performed on a workstation with an Intel Core i7 12th-generation processor, an NVIDIA RTX 3080 graphics card with 12 GB of VRAM, and 32 GB of RAM. Software included Windows 11 Pro, Python 3.10, TensorFlow 2.12, Keras 2.12, OpenCV 4.7, and packages for data visualization. The identical computational configuration was used for all models to provide equal opportunities for comparison of training and inference times. Model performance was assessed with measures of accuracy, sensitivity, specificity, precision, recall, F1 score, analysis of the confusion matrix, and determination of the area under the receiver operating characteristic curve. Overall accuracy reflects the general reliability of predictions, whereas sensitivity represents the reliability of positive predictions for the presence of cracks. Specificity provides a measure of the ability of the model to correctly identify the absence of cracks, and overall sensitivity provides a measure of detection efficiency for cracks. The ultimate goal for crack detection is to achieve high sensitivity and high specificity for the detection of cracks. However, the combination of very high sensitivity and low specificity results in excessive generation of false alarms. Conversely, high specificity with low sensitivity fails to detect existing structural defects.

4.4. Training Performance Analysis

The training and validation accuracy-loss curves were examined to evaluate convergence behavior and overfitting tendencies. Figure 5 presents the training and validation curves of the baseline CNN + InceptionV3 model.

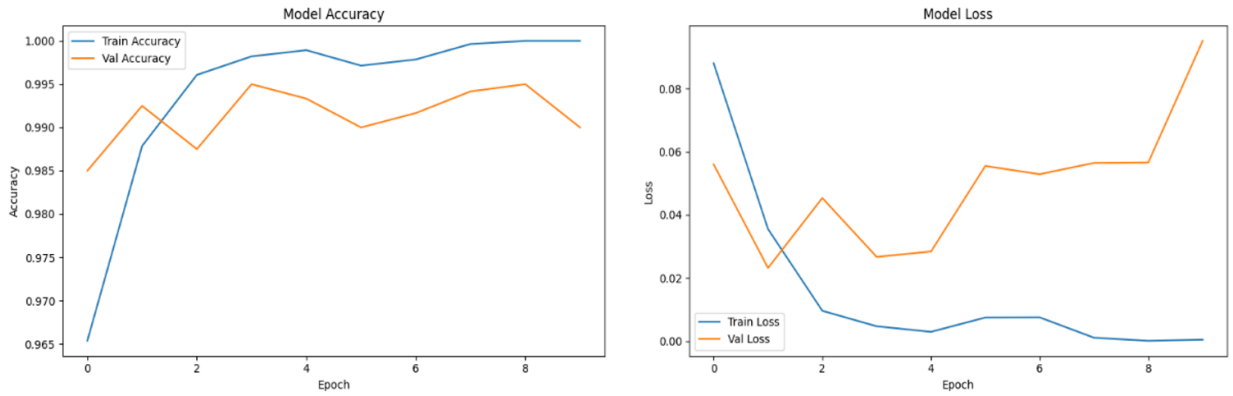


Figure 5. Training and validation accuracy-loss curves of the CNN + InceptionV3 baseline model.

© 2026 Conscientia Beam. All Rights Reserved.

As shown in Figure 5, the baseline model exhibited rapid convergence during the first few epochs and achieved stable validation accuracy of nearly 99%. Thus, transfer learning enabled efficient learning of crack-related features. However, the model was more sensitive to crack-like surface textures, as reflected by high recall and low precision. Training and validation curves for the CBAM + InceptionV3 model are shown in Figure 6.

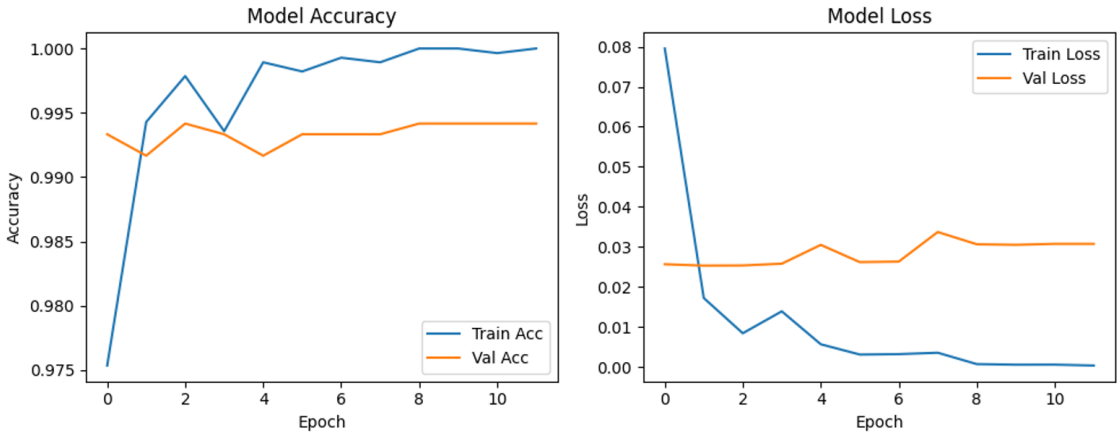


Figure 6. Training and validation accuracy-loss curves of the CBAM + InceptionV3 model.

Compared with the baseline model, the CBAM + InceptionV3 model demonstrated smoother validation behavior and fewer fluctuations in loss. This suggests that attention-guided feature refinement helped stabilize the learning process. The channel attention component emphasized the most informative feature maps, while spatial attention guided the model toward crack-relevant regions. Therefore, the improvement was not only in final accuracy but also in the consistency of learning.

4.5. Quantitative Performance Comparison

Table 3 presents the classification performance of the three models.

Table 3. Classification performance comparison of the evaluated models.

Model	Accuracy	Precision	Recall	Sensitivity	Specificity
CBAM + InceptionV3	99.17%	99.33%	99.00%	99.00%	99.33%
CNN + InceptionV3	99.00%	98.36%	99.67%	99.67%	98.33%
CBAM + VGG16	99.00%	98.84%	99.17%	99.17%	98.83%

Results of all three models in Table 3 demonstrate excellent classification accuracy. However, interpretation of precision, recall, and specificity showed greater differences among models. The baseline CNN with InceptionV3 provided the highest recall (99.67%) and thus achieved nearly complete detection of images with cracks. However, low precision and specificity resulted in substantial increases in the number of false-positive classifications for images without cracks and reflected major difficulties associated with surface crack detection. Contamination by stains, shadows, and coarse textures frequently produced an appearance similar to that of cracks.

The CBAM + InceptionV3 model provided the best overall balance with 99.17% accuracy, 99.33% precision, and 99.33% specificity. This demonstrates that CBAM reduced oversensitivity and improved reliability for crack predictions. Increased precision confirmed greater selectivity for detection of cracks, whereas increased specificity supported better discrimination of non-crack surfaces. Consequently, the major advantage of CBAM was not a small improvement in accuracy, but a more reliable limit between true cracks and patterns resembling cracks in the background. A comparison of classification performance for all models is shown in Figure 7.

© 2026 Conscientia Beam. All Rights Reserved.

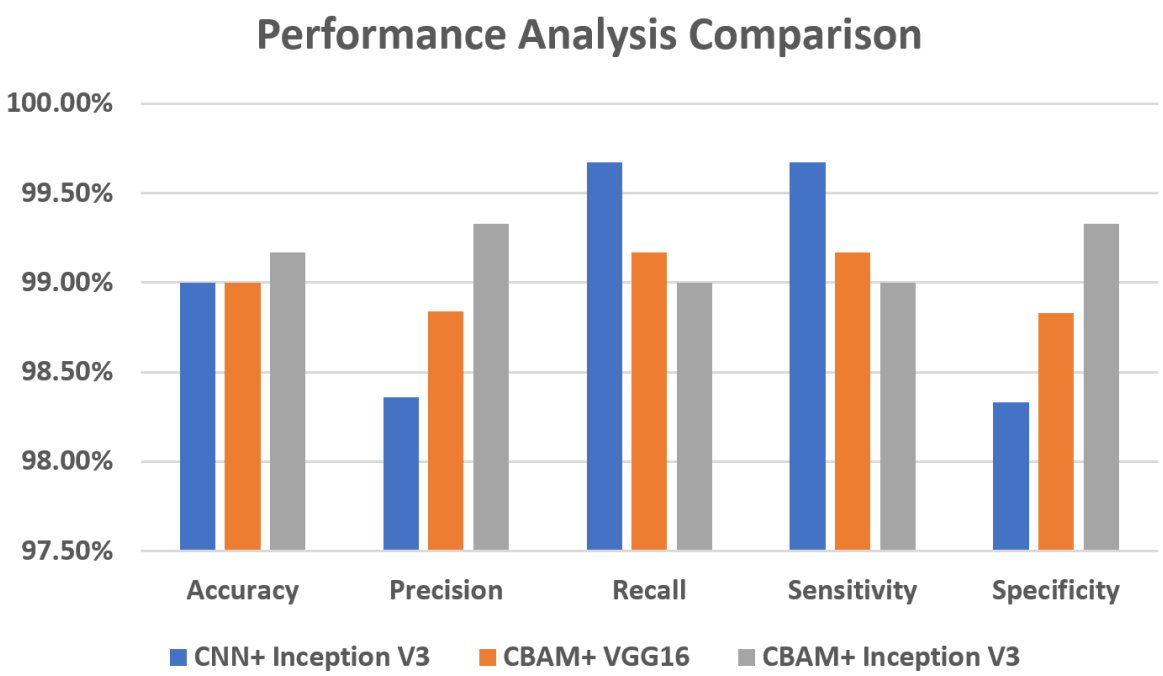


Figure 7. Comparative performance analysis of CNN + InceptionV3, CBAM + InceptionV3, and CBAM + VGG16 models.

As shown in Figure 7, the baseline model performs strongly in recall but has weaker precision and specificity. In contrast, the CBAM-enhanced models provide a better balance between crack detection and false-positive reduction. This supports the argument that attention-based refinement improves feature importance by directing the model toward structurally meaningful crack regions instead of responding to every crack-like texture.

4.6. Confusion Matrix Analysis

The confusion matrix of the CNN + InceptionV3 model is shown in Figure 8.

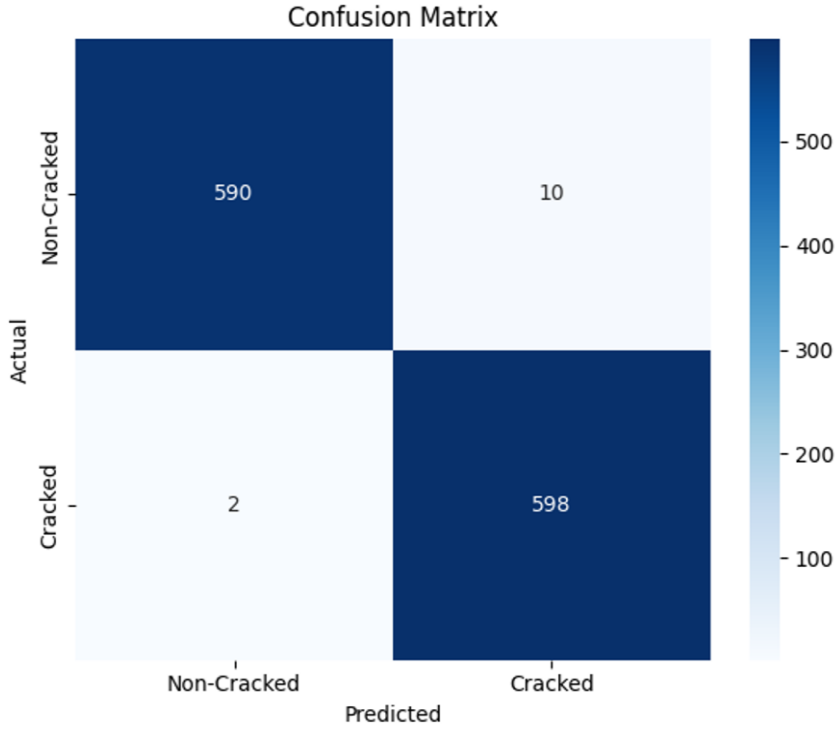


Figure 8. Confusion matrix of the CNN + InceptionV3 baseline model.

© 2026 Conscientia Beam. All Rights Reserved.

Figure 8 shows that the baseline model identifies nearly all crack samples correctly but misclassifies a relatively large number of non-crack samples as cracks. This corroborates the results in Table 1, for which the model provided high sensitivity but relatively low specificity and precision. In an operational inspection system, this would result in increased rates of unnecessary maintenance alerts and reduced user acceptance.

The confusion matrix of the CBAM + VGG16 model is shown in Figure 9.

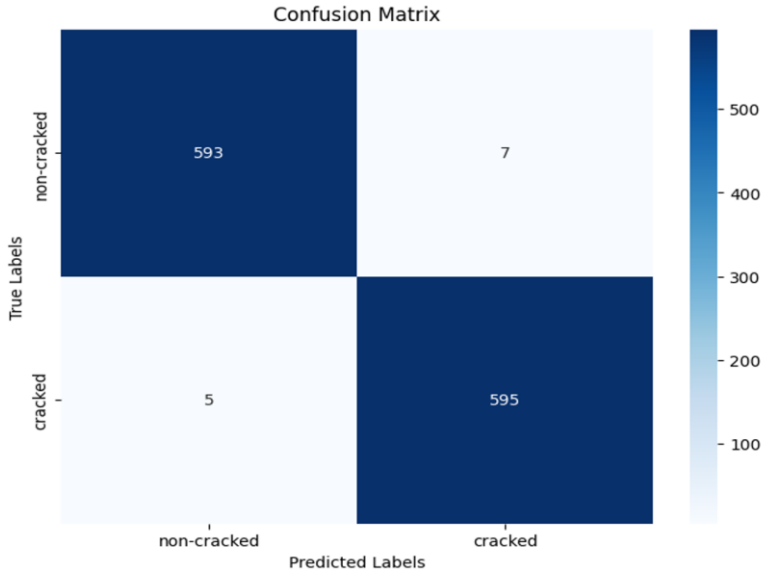


Figure 9. Confusion matrix of the CBAM + VGG16 model.

The CBAM + VGG16 model reduced false-positive predictions compared with the baseline and maintained competitive classification performance. Its lower model size and faster inference time make it suitable for resource-constrained deployment, although its feature extraction capability is comparatively less advanced than that of CBAM + InceptionV3. The confusion matrix of the CBAM + InceptionV3 model is shown in Figure 10.

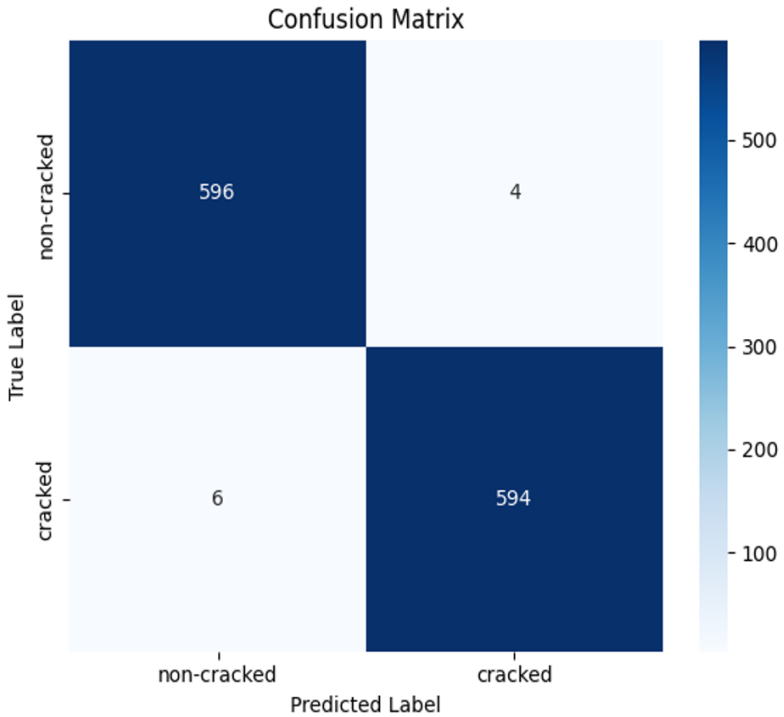


Figure 10. Confusion matrix of the CBAM + InceptionV3 model.

© 2026 Conscientia Beam. All Rights Reserved.

Figure 10 shows that CBAM + InceptionV3 achieved the most balanced classification behavior. The reduced number of false positives indicates that the attention module helped the model suppress misleading background activations. This is important in cement surface images because the defect region may occupy only a small part of the image, while the surrounding texture may dominate the feature representation if attention is not used.

4.7. Interpretation of Attention Effects and Feature Importance

The performance differences between baseline and CBAM-augmented models reflect the effects of channel and spatial attention. In a conventional CNN, all extracted feature maps contribute to classification based on learned convolutional filters. However, the importance of different feature maps varies with crack-detection requirements. Some feature maps reflect characteristics of surface texture, shadows, contaminants, or background patterns rather than features associated with cracks. Application of channel attention increases the relative importance of feature channels containing information relevant to crack detection and reduces the contribution of less informative feature channels.

Spatial attention enhances this process by emphasizing important image regions. Cracks to be detected are typically narrow, irregular, and sparsely distributed, and spatial attention enables the network to focus on local crack regions rather than applying a uniform representation across the entire surface. Consequently, integration of channel and spatial attention accounts for the improved performance of the CBAM+InceptionV3 network with regard to precision and specificity. The capacity of the network was not simply increased by the attention mechanism; rather, improved selectivity of extracted features was achieved.

This finding is consistent with recent attention-based crack detection methods that achieved improved discrimination of features by attention-driven learning [6, 8, 20, 21]. However, recent methods based on segmentation or transformers indicate that more complex models may be required for pixel-level localization of cracks or for estimation of crack widths [11, 12]. In contrast, we demonstrate here that for the task of binary classification of cracks, a simple attention mechanism such as CBAM provides an effective compromise among accuracy, reduction of false-positive classifications, and computational requirements without the need for pixel-level annotations.

4.8. Model Complexity and Deployment Readiness

Table 4 presents model complexity and deployment-related metrics.

Table 4. Model complexity and deployment-related metrics.

Model Architecture	Parameters (Millions)	Inference Time (ms)	Model Size (MB)
CNN + InceptionV3	23.8	41.2	91.7
CBAM + InceptionV3	24.5	43.6	95.3
CBAM + VGG16	17.9	36.1	71.5

As shown in Table 4, adding CBAM to InceptionV3 increased the number of parameters from 23.8 million to 24.5 million and increased inference time from 41.2 ms to 43.6 ms. This increase is small compared with the improvement in precision and specificity. Therefore, CBAM introduces only a modest computational overhead while improving decision reliability.

CBAM + VGG16 achieved the smallest model size of 71.5 MB and the fastest inference time of 36.1 ms. This indicates that CBAM + VGG16 may be useful for low-resource or edge-based monitoring systems. However, CBAM + InceptionV3 remains more suitable for accuracy-critical inspection scenarios, where reducing false positives is more important than minimizing model size.

© 2026 Conscientia Beam. All Rights Reserved.

4.9. Statistical Validation

To determine whether the observed improvement was statistically meaningful, a paired two-tailed t-test and McNemar's test were applied between the CNN + InceptionV3 baseline and the CBAM + InceptionV3 model. The tests showed that the increase in precision and reduction in false positives achieved by the attention-augmented model were statistically significant at a 95% confidence level, with $p < 0.05$.

This statistical validation strengthens the interpretation of the results because the improvement is not discussed only as a numerical difference in accuracy. Instead, the analysis confirms that CBAM contributes to more reliable classification behavior, particularly by reducing false alarms. This is important because many crack detection studies report very high accuracy on benchmark datasets but do not test whether differences between models are statistically meaningful.

4.10. Cross-Validation Analysis for Generalization Reliability

Five-fold cross-validation was used to test the stability of the high classification performance across different data partitions. The data were split into five independent subsets. In each cycle, four subsets were used for training and one for testing. As a result, each sample was tested exactly once.

Table 5. Five-fold cross-validation performance of the CNN + InceptionV3 model.

Fold	Accuracy	Precision	Recall	Specificity	F1	AUC
1	0.985	0.985	0.985	0.985	0.985	0.999
2	0.990	0.990	0.990	0.989	0.990	0.998
3	0.988	1.000	0.976	1.000	0.988	0.998
4	0.990	1.000	0.981	1.000	0.990	0.999
5	0.992	0.995	0.989	0.995	0.992	0.999
Average	0.989	0.994	0.984	0.994	0.989	0.999

Table 5 shows that accuracy remained between 98.52% and 99.26%, with an average accuracy of 98.91%. The mean AUC of 0.9991 indicates strong class separability between cracked and non-cracked images. The small fold-to-fold variation suggests that the model is not simply memorizing a particular train-test split.

Figure 11 presents the cross-fold performance stability of the model.

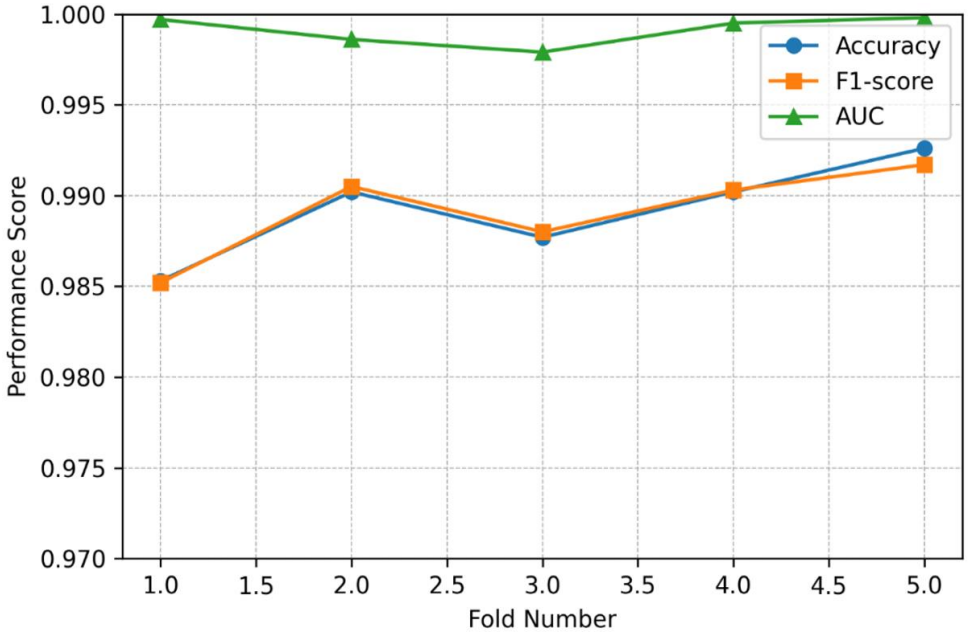


Figure 11. Cross-fold performance stability analysis using five-fold cross-validation.

© 2026 Conscientia Beam. All Rights Reserved.

As shown in Figure 11, the performance variation across folds is very small. This supports the reliability of the reported results and helps address the common concern that very high crack detection accuracy may be caused by data leakage or benchmark saturation. However, the dataset remains a controlled benchmark dataset; therefore, future validation on more diverse field images would further strengthen practical generalization.

4.11. Hyperparameter Optimization and Training Efficiency

Table 6 presents the performance obtained during hyperparameter optimization. These results were obtained with the procedure used to optimize the behavior of the optimizer, dropout regularization and training stability. As a result, they should not be interpreted as an alternative to previously reported independent test-set results; rather, they provide a means for comparing the relative robustness and efficiency of different model configurations during optimization. All results are given to three decimal places for consistency.

Table 6. Performance comparison after hyperparameter tuning.

Model	Accuracy	Precision	Recall	F1	AUC	Training Time (Min)
CBAM + InceptionV3	0.983	0.983	0.983	0.983	0.995	11.202
CBAM + VGG16	0.981	0.986	0.976	0.981	0.996	25.558
CNN + InceptionV3	0.983	0.986	0.979	0.983	0.997	10.864

The accuracy differences among the three models were small, as expected for a well-established binary crack classification benchmark with high class separability. However, the influence of model configuration was reflected in the trade-off between precision, recall, AUC and training time. The fastest training time (10.864 min) was achieved with the baseline CNN + InceptionV3 architecture, whereas the addition of CBAM increased training time by only 0.34 min. This represents a minimal additional computational burden for training.

CBAM + VGG16 required the longest training time, at 25.558 minutes, despite its smaller model size. This may be attributed to the sequential convolutional structure of VGG16, which is less computationally optimized than InceptionV3 for multi-scale feature extraction. Therefore, model size alone does not determine training efficiency. Architecture design, convolutional block arrangement, feature extraction strategy, and optimization behavior also influence the training process. Overall, the hyperparameter-tuning results support using CBAM + InceptionV3 as a balanced configuration, as it provides competitive validation performance with only a minor increase in training time compared with the baseline model.

4.12. Ablation Study: Impact of Attention Mechanism

An ablation study was performed to determine the individual and combined effects of channel and spatial attention. As shown in Table 7, the baseline CNN + InceptionV3 model was incrementally improved with channel attention alone, spatial attention alone and with full CBAM integration.

Table 7. Ablation analysis of CBAM components.

Model Variant	Channel Attention	Spatial Attention	Accuracy	Precision	Recall	F1-Score	AUC
CNN + InceptionV3 Baseline	✗	✗	0.983	0.986	0.979	0.983	0.997
Channel Attention Only	✓	✗	0.984	0.987	0.980	0.984	0.998
Spatial Attention Only	✗	✓	0.985	0.988	0.981	0.984	0.998
CBAM: Channel + Spatial	✓	✓	0.992	0.993	0.990	0.992	0.999

Our ablation results confirm that both channel and spatial attention improve classification performance. Channel attention improves feature discrimination by suppressing uninformative feature maps, whereas spatial attention increases selectivity for regions of interest by suppressing non-relevant regions in images of cement surfaces. Spatial

© 2026 Conscientia Beam. All Rights Reserved.

attention alone yielded slightly greater accuracy and precision than channel attention alone, thereby demonstrating the importance of localizing features relevant to crack detection to distinguish cracks from variations in background texture.

The overall performance of the full CBAM configuration (0.992 accuracy, 0.993 precision, 0.990 recall, 0.992 F1 score, and 0.999 AUC) represents only a modest improvement over the baseline model. However, because the present dataset represents a highly separable situation in which baseline performance is already excellent, the results confirm the utility of CBAM for reducing the rate of false-positive predictions and for improving the consistency of inspection decisions, rather than for achieving large increases in overall accuracy.

4.13. Benchmark Comparison and Discussion with Recent Studies

Recent crack detection studies have reported high accuracy using CNNs, YOLO-based models, transformer-enhanced networks, and segmentation frameworks [11-13, 20, 21, 24, 25]. These studies confirm that deep learning methods are highly effective for automated crack inspection. However, many of them focus on either detection accuracy, segmentation performance, or model innovation without providing a controlled comparison of attention modules across CNN backbones.

The findings of the present study are consistent with attention-based studies that show improved feature selectivity and robustness in crack detection [6, 8, 20]. The results also support the argument that attention mechanisms are useful when the background contains crack-like textures. However, some segmentation-based works argue that image-level classification alone is insufficient for practical inspection because maintenance decisions may require crack width, crack length, or pixel-level localization [11, 12]. This is a valid limitation of the present study, as the proposed models perform binary classification rather than pixel-level segmentation.

Compared with lightweight and edge-focused studies [14, 15, 22, 23], the present work provides a balanced interpretation. CBAM + VGG16 offers a smaller model size and faster inference, while CBAM + InceptionV3 provides stronger classification reliability. Therefore, the preferred model depends on deployment requirements. For mobile or low-power devices, CBAM + VGG16 may be more practical. For inspection systems where false-positive reduction and decision reliability are critical, CBAM + InceptionV3 is more suitable.

Overall, the results show that CBAM is necessary not because the dataset cannot be solved by standard CNNs, but because attention improves the quality of model decisions. In infrastructure monitoring, a model with slightly higher accuracy but many false alarms may be less useful than a model that provides balanced precision, specificity, and stable predictions. Thus, the proposed attention-enhanced framework contributes to more reliable automated crack detection.

4.14. Experimental Findings

The experimental results show that attention-enhanced architectures improve the reliability of cement surface crack detection. The CBAM+InceptionV3 model achieved the highest overall performance, with 99.17% accuracy and 99.33% precision. It produced fewer false positives and thus demonstrated reduced sensitivity to crack-like textures. In contrast, the baseline CNN+InceptionV3 model achieved higher recall at the expense of greater false-positive rates. The resulting CBAM+VGG16 model achieved competitive performance with reduced model complexity and faster inference times and thus was well suited for applications with limited computational resources. Results of our statistical validation, cross-validation analyses, comparisons of model complexity, and an ablation study confirmed that improvements attributable to CBAM were not merely a consequence of architectural complexity but largely reflected improved importance of channel-based features and greater spatial discrimination. Consequently, our results support the application of attention-based mechanisms for reliable and computationally efficient structural health-monitoring systems.

© 2026 Conscientia Beam. All Rights Reserved.

5. DISCUSSION AND LIMITATIONS

The results confirm that attention-augmented CNN architectures significantly improve the reliability of automated crack detection on cement surfaces, especially by reducing false-positive predictions and increasing precision. However, all models achieved excellent performance, with the best overall balance achieved by the CBAM + InceptionV3 architecture. This configuration produced high values for accuracy, precision, recall, and specificity, and confirmed the utility of CBAM for refining the importance of features and distinguishing true crack regions from visually similar surface textures such as stains, shadows, coarse textures, and background-texture irregularities.

The baseline CNN + InceptionV3 model achieved the highest recall and thus was highly sensitive to the detection of crack samples. However, its relatively low precision and specificity reflected a higher likelihood of misclassifying non-crack images as cracked. In practical applications to infrastructure inspection, this resulted in increased rates of unnecessary maintenance alerts and reduced reliability of automated inspection systems. In contrast, the performance of the CBAM-enhanced models was superior with respect to both crack detection and reduction of false-positive rates. This was achieved by the sequential application of channel and spatial attention to produce more selective and meaningful representations of image features. Channel attention facilitated the selection of the most informative feature maps, whereas spatial attention increased sensitivity to regions of the image corresponding to cracks. As a result, CBAM enabled more effective learning of meaningful image features.

The results are consistent with recent attention-based and hybrid crack detection studies that report improved feature discrimination and robustness through attention-guided learning [6, 8, 20, 21]. However, some recent segmentation-based and transformer-based studies suggest that classification alone may not be sufficient for detailed structural assessment because practical maintenance decisions may also require crack width, crack length, severity level, and pixel-level localization [11, 12]. This highlights an important distinction in the present work. The proposed framework is designed for image-level binary crack classification, not for crack segmentation or severity quantification. Therefore, its contribution lies mainly in improving classification reliability and deployment readiness rather than providing pixel-level structural diagnosis.

Another important observation is that the dataset used in this study is a well-known benchmark dataset, and several previous studies have already reported very high accuracy on similar crack detection datasets. Therefore, the reported high accuracy should not be interpreted as the only contribution of this work. Instead, the main value of the present study lies in its controlled comparison of baseline and CBAM-enhanced architectures, cross-validation-based reliability assessment, statistical validation, ablation analysis, and deployment-oriented comparison using model size, inference time, and training efficiency. These additional evaluations provide a more comprehensive understanding of how attention mechanisms affect model behavior beyond accuracy alone.

Despite its strengths, the study is subject to certain limitations. First, our dataset includes only two classes of crack condition (cracked and non-cracked) and does not incorporate multiple classes of crack type, such as transverse, longitudinal, diagonal, map, or block cracks. Second, we did not estimate the severity, width, depth, or length of cracks, which are important parameters for planning structural maintenance. Third, our results are based on image-level classifications and thus did not include pixel-level segmentation or precise localization of cracks. Fourth, the use of five-fold cross-validation minimized bias associated with data partitioning, but continued reliance on controlled benchmark images limited the degree of generalizability achieved. Additional validation with field images obtained under different environmental and structural conditions would have further increased the degree of generalizability. Finally, we did not incorporate methods for visualization of model outputs because our objective was to evaluate architectural performance, reliability of classifications, and feasibility of implementation, rather than to obtain a detailed interpretation of model behavior. However, application of techniques such as Grad-CAM in future work would provide additional information regarding whether the model focuses on actual regions of crack damage.

Overall, the discussion supports the view that CBAM is valuable not because it substantially improves accuracy for an already well-established benchmark problem, but because it increases the quality and reliability of model

© 2026 Conscientia Beam. All Rights Reserved.

decisions. For practical structural health monitoring, a model with reduced rates of false positives and consistent performance across validation conditions is of greater value than a model that provides only high rates of accuracy. Future efforts should expand application of the proposed methodology to multi-class crack classification, estimation of crack widths and severity, pixel-level segmentation, methods for interpretation based on Grad-CAM, and validation with large data sets representing actual conditions of infrastructure systems.

6. CONCLUSION AND FUTURE SCOPE

This work reports on a controlled comparative assessment of attention-enhanced deep learning architectures for automated detection of cement surface cracks. The primary aim was to determine whether incorporation of the convolutional block attention module into CNN-based architectures would improve classification performance, reduce false positives, and enable practical applications for structural health monitoring. Results for three model configurations (i.e., CNN + InceptionV3, CBAM + InceptionV3, and CBAM + VGG16) were obtained using an identical experimental protocol, dataset, preprocessing methodology, training conditions, and criteria for evaluation of model performance.

The experimental findings show that the CBAM + InceptionV3 model achieved the best overall performance, with 99.17% accuracy and 99.33% precision. The results confirm that CBAM improves the model's ability to focus on crack-relevant features while suppressing misleading background textures, stains, and surface irregularities. Although the baseline CNN + InceptionV3 model achieved high recall, it produced comparatively more false positives. In contrast, the CBAM-enhanced models provided a better balance between precision, recall, specificity, and decision reliability. The CBAM + VGG16 model also achieved competitive performance with lower computational complexity, indicating its potential for resource-constrained deployment.

The practical and policy implications of this work are that attention-enhanced automated systems for crack detection can assist agencies responsible for infrastructure monitoring, civil engineers, and personnel engaged in maintenance planning by providing faster, more consistent, and less subjective screening for cracks. These systems will facilitate prioritization of inspection activities, reduce reliance on manual visual assessments, support planning for preventive maintenance, and optimize allocation of resources for public infrastructure safety.

Future work will expand the application of our approach to multiclass classification of crack types, estimation of crack severity, measurement of crack widths, and pixel-based segmentation of cracks. In addition, the application of methods for explainable artificial intelligence will increase the capacity for interpretation. Finally, our results provide support for further development and application of lightweight convolutional neural networks, models based on transformers, and architectures that permit real-time operation for monitoring of infrastructure conditions.

Funding: This study received no specific financial support.
Institutional Review Board Statement: Not applicable.
Transparency: The authors state that the manuscript is honest, truthful, and transparent, that no key aspects of the investigation have been omitted, and that any differences from the study as planned have been clarified. This study followed all writing ethics.
Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: Both authors contributed equally to the conception and design of the study. Both authors have read and agreed to the published version of the manuscript.
Disclosure of AI Use: The author(s) used OpenAI's ChatGPT to improve language, grammar, and manuscript readability. All generated outputs were carefully reviewed and verified by the authors.

REFERENCES

[1] R. Fu, M. Cao, D. Novák, X. Qian, and N. F. Alkayem, "Extended efficient convolutional neural network for concrete crack detection with illustrated merits," *Automation in Construction*, vol. 156, p. 105098, 2023. https://doi.org/10.1016/j.autcon.2023.105098

[2] H. Kaveh and R. Alhajj, "Recent advances in crack detection technologies for structures: A survey of 2022-2023 literature," *Frontiers in Built Environment*, vol. 10, p. 1321634, 2024. https://doi.org/10.3389/fbuil.2024.1321634

© 2026 Conscientia Beam. All Rights Reserved.

[3] Q. Yuan, Y. Shi, and M. Li, "A review of computer vision-based crack detection methods in civil infrastructure: Progress and challenges," *Remote Sensing*, vol. 16, no. 16, p. 2910, 2024. https://doi.org/10.3390/rs16162910

[4] V. P. Golding, Z. Gharineiat, H. S. Munawar, and F. Ullah, "Crack detection in concrete structures using deep learning," *Sustainability*, vol. 14, no. 13, p. 8117, 2022. https://doi.org/10.3390/su14138117

[5] 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, 2020. https://doi.org/10.3151/jact.18.493

[6] H. Yuan, T. Jin, and X. Ye, "Modification and evaluation of attention-based deep neural network for structural crack detection," *Sensors*, vol. 23, no. 14, p. 6295, 2023. https://doi.org/10.3390/s23146295

[7] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," presented at the Proceedings of the European Conference on Computer Vision, 2018.

[8] B. Ozdemir, F. Sermet, and I. Pacal, "Attention-enhanced ConvNeXt for accurate, efficient, and interpretable crack detection," *Expert Systems with Applications*, vol. 296, p. 129165, 2026. https://doi.org/10.1016/j.eswa.2025.129165

[9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014. https://doi.org/10.48550/arXiv.1409.1556

[10] S. Dorafshan, R. J. Thomas, and M. Maguire, "Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete," *Construction and Building Materials*, vol. 186, pp. 1031-1045, 2018. https://doi.org/10.1016/j.conbuildmat.2018.08.011

[11] C. Xiang, W. Wang, L. Deng, P. Shi, and X. Kong, "Crack detection algorithm for concrete structures based on super-resolution reconstruction and segmentation network," *Automation in Construction*, vol. 140, p. 104346, 2022. https://doi.org/10.1016/j.autcon.2022.104346

[12] D. Dais, I. E. Bal, E. Smyrou, and V. Sarhosis, "Automatic crack classification and segmentation on masonry surfaces using convolutional neural networks and transfer learning," *Automation in Construction*, vol. 125, p. 103606, 2021. https://doi.org/10.1016/j.autcon.2021.103606

[13] X. Dong, Y. Liu, and J. Dai, "Concrete surface crack detection algorithm based on improved YOLOv8," *Sensors*, vol. 24, no. 16, p. 5252, 2024. https://doi.org/10.3390/s24165252

[14] 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, p. 115, 2021. https://doi.org/10.3390/infrastructures6080115

[15] A. Raza, F. Hanif, and H. A. Mohammed, "Efficient crack and surface-type recognition via CNN-block development mechanism and edge profiling," *Scientific Reports*, vol. 15, no. 1, p. 40073, 2025. https://doi.org/10.1038/s41598-025-25956-8

[16] Y. Zhang, G. Li, and R. Shen, "Computer vision for infrastructure defect detection: Methods and trends," *Automation in Construction*, vol. 183, p. 106795, 2026. https://doi.org/10.1016/j.autcon.2026.106795

[17] Y. Choi, H. W. Park, Y. Mi, and S. Song, "Crack detection and analysis of concrete structures based on neural network and clustering," *Sensors*, vol. 24, no. 6, p. 1725, 2024. https://doi.org/10.3390/s24061725

[18] M. Shahin, F. F. Chen, M. Maghanaki, A. Hosseinzadeh, N. Zand, and H. Khodadadi Koodiani, "Improving the concrete crack detection process via a hybrid visual transformer algorithm," *Sensors*, vol. 24, no. 10, p. 3247, 2024. https://doi.org/10.3390/s24103247

[19] L. Ali, F. Alnajjar, H. A. Jassmi, M. Gocho, 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, p. 1688, 2021. https://doi.org/10.3390/s21051688

[20] B. A. Kyem, J. K. Asamoah, E. Denteh, A. Danyo, and A. Aboah, "Self-supervised multi-scale transformer with Attention-Guided Fusion for efficient crack detection," *Automation in Construction*, vol. 181, p. 106591, 2026. https://doi.org/10.1016/j.autcon.2025.106591

© 2026 Conscientia Beam. All Rights Reserved.

[21] D. P. Yadav, B. Sharma, S. Chauhan, and I. B. Dhaou, "Bridging convolutional neural networks and transformers for efficient crack detection in concrete building structures," *Sensors*, vol. 24, no. 13, p. 4257, 2024. https://doi.org/10.3390/s24134257

[22] M. Yu, W. Chen, and J. Hou, "A lightweight network for concrete crack detection in complex environments based on cloud-edge collaboration," *Expert Systems with Applications*, vol. 299, p. 130204, 2026. https://doi.org/10.1016/j.eswa.2025.130204

[23] N. Isong, "Building efficient lightweight CNN models," *arXiv preprint arXiv:2501.15547*, 2025. https://doi.org/10.48550/arXiv.2501.15547

[24] S. S. R. Krishnan *et al.*, "Comparative analysis of deep learning models for crack detection in buildings," *Scientific Reports*, vol. 15, no. 1, p. 2125, 2025. https://doi.org/10.1038/s41598-025-85983-3

[25] S. Matarneh, F. Elghaish, F. P. Rahimian, E. Abdellatef, and S. Abrishami, "Evaluation and optimisation of pre-trained CNN models for asphalt pavement crack detection and classification," *Automation in Construction*, vol. 160, p. 105297, 2024. https://doi.org/10.1016/j.autcon.2024.105297

Views and opinions expressed in this article are the views and opinions of the author(s), Review of Computer Engineering Research shall not be responsible or answerable for any loss, damage or liability etc. caused in relation to/arising out of the use of the content.

© 2026 Conscientia Beam. All Rights Reserved.