“Bodh”, BPIT’s International Journal of Technology & Management
BBIJTM, ISSN: 2454-8421, Volume 11, 2025

Automatic Concrete Crack Detection and Analysis by AI Model

Shivam Sharma[1], Tanima Ghosh[2][0000-1111-2222-3333] , Pavika Sharma[3][1111-2222-3333-4444] and Preeti Singh[4][1111-2222-3333-4444]

[1,2,3,4] Department of Electronics and Communication Engineering, Bhagwan Parshuram Institute of Technology, Sector-17, Rohini, Delhi-89, (Affiliated to GGSIPU)
tanimaghosh@bpitindia.com

Abstract. The AI based techniques act as groundbreaking breakthrough for Concrete Crack Detection and Analysis. The work presents design of a smart computational framework capable of automated identification, classification and further exhaustive analysis of the occurring defects in concrete structures. To enhance the accuracy and reliability of the defect detection system, we use image processing integrated with deep learning models. Res-Net 50 based deep convolutional neural network is employed for the feature extraction utilizing residual learning methods and skip connections to allow the network to learn hierarchical representation of the crack patterns and the irregularities present in surface. We extract the feature vectors of concrete surfaces to capture local texture pattern and global structural characteristics. Following feature extraction, several classification strategies are implemented to find the most effective model for concrete defect identification. A Support Vector Machine (SVM) classifier is initially implemented for handling the high-dimensional feature spaces and complex decision boundaries. After that hybrid combinations are made by integrating SVM with Random Forest (RF), SVM with Gradient Boosting (GB) and SVM with RF and GB. The experimental results reveals that the ensemble combination SVM with RF and GB achieved the best classification accuracy of 99.81%. In addition to classification, YOLOv7 object detection algorithm is used for precise localization of the cracks.

Keywords: SVM, Gradient Boost, Random Forest, YOLOv7, AI, ML

1 Introduction

Automatic concrete crack detection becomes very crucial and essential now a days [1]. Traditional methods relied on manual visual inspection done by trained professionals.

Although these practices have served as the basis for structural health monitoring, they have many limitations. Manual monitoring procedures are often labour-intensive, time-consuming, and dependent on human judgement, and can introduce non reliable result. These limitations demand the need of automatic monitoring of concrete structure for automatic detection and analysis of its faults.

A detailed analysis of automated crack detection based on the existing models are summarized in table 1.

Table 1. Comparative Analysis of Advanced Crack Detection Technologies

Reference No.	Authors/ Year	Technology used	Key Advantages	Limitations
[1]	Yu et al., 2022	YOLOv5s Deep-Learning Architecture	Enhanced real-time detection with optimized computational efficiency	Insufficient sensitivity for micro-crack detection
[2]	Nyathi et al. 2024	Custom deep CNN and U-Net models	Quantify crack width in millimeters	Require large size data set
[3]	Wang et al., 2025	Improves MobileNetV3 and channel attention	Light weight model with better feature representation for crack classification	Elevated computational complexity
[4]	Khan et al., 2023	Systematic Literature Review	Comprehensive survey of deep learning methodologies for concrete defect detection	Review paper—no empirical implementation
[5]	C.P.P. et al., 2024	YOLOv Architecture	Reduced computational resource requirements	Compromised real-time detection accuracy
[6]	Harb et al. 2025	Swing UNETR architecture	Outperformed mono-temporal U-Net	-
[7]	Shibano et al., 2024	CNN with Infrared Imaging	Capability for subsurface defect detection	Limited efficacy for microscopic crack identification
[8]	Choi et al., 2024	CNN with Clustering Algorithms	Enhanced robustness for complex structural geometris.	Increased system architectural complexity
[9]	Chen et al., 2025	U-Net Pixel-level Segmentation	High-precision crack morphology quantification	Significant computational resource demands

[10]	Shi et al., 2025	Kolmogorov-Arnold Network (KAN)	Superior performance relative to Vision Transformer variants	Elevated inference latency and hardware requirements

Despite substantial progress in automated crack detection technologies, significant challenges persist in achieving reliable defect identification under adverse imaging conditions. specifically, the presence of complex background interference including surface contamination, weathering-induced discoloration, extraneous markings, and variable illumination conditions continues to compromise detection accuracy and generate false-positive classification.

To overcome these limitations, this research proposes a novel intelligent framework for structural defect detection and analysis that integrates a multi-stage pre-processing process with advanced deep learning techniques. The proposed approach utilizes Res-Net-50 to extract features from the pre-processed images. These features are used for defect classification through a comparative evaluation of multiple ML classifier that include Support Vector machine (SVM), hybrid combination of SVM-Random Forest (RF), SIM-Gradient-Boost (GB), and an ensemble framework that integrates SVM-RF-GB. This multi-classifier evaluation enables a comprehensive assessment of the most effective strategy for reliable structural defect identification and analysis. The system integrates YOLOv7 object detection architecture to provide precise spatial localization and geometric characterization of identified crack regions.

2 Methodology

The overall architecture of the proposed model is illustrated in figure 2.1. This methodology consists of five major stages:

(i) Dataset acquisition
(ii) Image pre-processing
(iii) Hierarchical feature extraction using ResNzet-50
(iv) Ensemble classification using hybrid ML models
(v) Spatial defect localization using YoLov7 architecture.

These stages collectively enable accurate detection, classification and localization of structural cracks in complex environmental conditions. These stages collectively enable accurate detection, classification, and localization of structural cracks in complex environmental conditions.

The flow diagram of the proposed model is shown in figure 2 below.

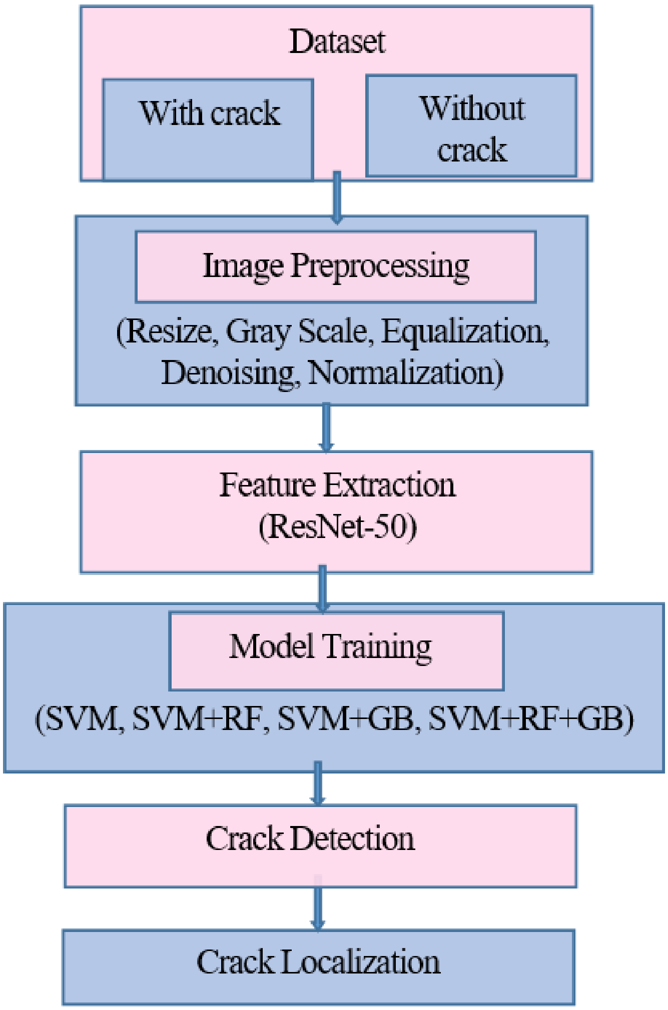

Fig. 2 Architectural flow diagram of the proposed multi-stage crack detection framework

2.1 Dataset Acquisition: The proposed system utilizes a publicly available concrete crack image repository sourced from the Kaggle platform [11].

2.2 Pre-processing: The following steps are involved in the pre-processing stage:

Grayscale Conversion: The initial pre-processing stage involves the RGB to Gray scale image conversion. The redundancy in color information can be reduced by this stage.

Contrast Enhancement Using Contrast Limited Adaptive Histogram Equalization (CLAHE) is utilized to enhance the local contrast of the images.

Noise reduction via Gaussian Filtering Gaussian blur filtering is used to eliminate the high frequency components of an image and to improve the surface irregularities and environmental interference.

Batch Normalization Batch normalization is incorporated to standardize input data distribution across training batches, accelerating convergence dynamics and enhance model training stability.

Interpolation for Resolution Standardization Bi-linear interpolation algorithms are employed to standardize image dimensions, ensuring uniform input geometry compatible with subsequent feature extraction architecture.

Pre-processing Quality Metrics The efficiency of the pre-processing pipeline is quantitatively assessed using established image quality metrics, specifically Peak Signal to Noise Ratio (PSNR) and Mean Squared Error (MSE).

The following formulas have been used to calculate these metrics:

$$(PSNR)_{DB} = (10 \times log_{10}(Peak\ value)^2)/MSE \qquad (2.2)$$

Where, MSE is represented as:

$$MSE = \frac{1}{xy}\left(I^{"}(x,y) - I(x,y)\right)^2 \qquad (2.3)$$

For 8-bit Gray-scale images the Peak value is 255.

2.3 Deep Learning-Based Feature Extraction: The pre-processed images are propagated through the ResNet50 architecture, undergoing successive convolution that progressively abstract low-level edge and texture information into high-level semantic representations. Early layers capture fundamental visual primitives including edge corners, and simple textures. Intermediate layers synthesize these primitives into more complex structural patterns, while deeper layers encode abstract crack characteristics invariant to minor geometric and photometric variations. The final convolutional layer out-puts a high dimensional feature tensor encapsulate discriminative crack related information. This feature representation undergoes global average pooling to generate a fixed-length feature vector suitable for subsequent classification operations, effectively condensing spatial feature maps into a compact descriptor while maintaining representational capacity.

2.4 Ensemble Classification Framework

Support Vector Machine (SVM) Baseline: SVM constitute the foundational classification architecture employed in this research.

SVM-Random Forest (RF) Hybrid Architecture: This cascaded architecture leverages the geometric margin optimization characteristics of SVM while incorporating the variance reduction properties of RF across diverse test conditions.

SVM-Gradient Boost (GB) Hybrid Architecture: The SVM-GB architecture corrects mis-classifications by sequential error correction mechanisms.

SVM-RF-GB Hybrid Architecture: The 4th classification strategy integrates 3 different classification algorithms for better classification result and achieved the optimal classification accuracy of 99.81%.

2.5 Spatial Localization Using YOLOv7: The spatial localization of cracks are done through YOLOv7 (You Only Look Once version 7).

3 Experimental Results and Performance Analysis

3.1 Pre-processing Optimization and Quality Assessment

Experimental Configuration: The pre-processing optimization phase used the dataset comprised 40,000 images with quality metrics computed as averaged values across the entire corpus to ensure statistical robustness and generalization validity. Figure 3 illustrates a representative input image from the dataset.

Fig. 3 Representative sample input image demonstrating typical concrete surface characteristics

Parameter Optimization Results: After pre-processing, the PSNR value achieved is 35.02 dB with CLAHE clip-limit 1.0 and grid size 8X8, Gaussian blur (5X5, and standard deviation =0), and with sharpening kernel ([0, -0.5, 0], [-0.5, 0, -0.5], [0, -0.5, 0].

3.2 Classification Performance Evaluation

Experimental Methodology: After the optimized pre-processing stage, feature vectors from ResNet-50 were employed to train four different classification models. The effectiveness of each model was evaluated through confusion matrix analysis, along with performance metrics such as accuracy, mis-classification rate, precision, recall and specificity.

Confusion Matrix Analysis: Table 3.4 displays the confusion matrix (CM) for the four classification models, outlining the number of true positive (TP), true negative (TN), false positive(FP), false negative (FN). The matrix demonstrates a consistent enhancement in performance across the ensemble models, showing a gradual decrease in both FP and FN values as the approach advances from the baseline SVM to SVM combined with RF, then to SVM with GB, and finally to the integrated SVM+RF+GB framework.

3.3 Performance Metrics Formulation

Classification performance is quantitatively characterized using the following standard evaluation metrics:

Accuracy: Represents the proportion of correct predictions and is represented by equation 3.1.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{3.1}$$

Mis-classification Rate: Quantifies the proportion of erroneous predictions and is represented by equation 3.2.

$$Mis\ Classification\ Rate = \frac{FP+FN}{TP+TN+FP+FN} \tag{3.2}$$

Precision: Measures the proportion of positive predictions that are correctly identified and is represented by equation 3.3.

$$Precision = \frac{TP}{TP+FP} \tag{3.3}$$

Recall (Sensitivity): Quantifies the proportion of actual positives correctly identified and is represented by equation 3.4.

$$Recall = \frac{TP}{TP+FN} \tag{3.4}$$

Specificity: Represents the proportion of actual negatives correctly identified and is represented by equation 3.5.

$$Specificity = \frac{TN}{TN+FP} \tag{3.5}$$

3.4 Comparative Classification Results

Table 2 presents comprehensive performance metrics for the four classification architectures

Table 2. Comparative Classification Performance Across Different Architectures

Model	TP	TN	FP	FN	Accuracy	Misclassification	Precision	Recall	Specificity
SVM	5987	5926	82	5	0.9927 (99.27%)	0.0073 (0.73%)	0.9865 (98.65 %)	0.9992 (99.92%)	0.9864 (98.64%)
SVM+RF	6036	5915	33	16	0.9959 (99.59%)	0.0041 (0.41%)	0.9946 (99.46 %)	0.9974 (99.74%)	0.9945 (99.45%)
SVM+GB	6053	6015	16	16	0.9973 (99.73%)	0.0027 (0.27%)	0.9974 (99.74 %)	0.9974 (99.74%)	0.9973 (99.73%)
SVM+RF+GB	5980	5997	11	12	0.9981 (99.81%)	0.0019 (0.19%)	0.9982 (99.82 %)	0.99799 (99.80%)	0.99817 (99.82%)

3.5 Performance Analysis and Discussion

The integrated ensemble of SVM, RF and Gradient Boosting yield the highest accuracy of 99.81%. This final hybrid model produces very few mis-classifications and demonstrates superior overall classification performance.

Spatial Localization Performance Using YOLOv7

Training and Validation Loss Analysis: The YOLOv7- based object detection framework was employed to identify the localize cracks in images that were previously labeled as positive by the SVM+RF+GB framework. Table 3 summarizes the final loss values obtained at the end of the training process, reflecting effective convergence and proper optimization of the detection network.

Table 3. YOLOv7 Training and Validation Loss Components

Loss Component	Training	Validation
Object Loss	0.005620	0.004809
Box Regression Loss	0.020800	0.024933
Classification Loss	0.000020	0.000030

3.6 Comparative analysis with the existing models

The proposed SVM+RF+GB system achieved 99.81% detection accuracy, with precision and Recall value 92.94% and 92.94% respectively. But the existing CNN model [1] achieved the precision and Recall value 72.75% and 80.82% respectively. So, we can say that our proposed model gives better performance in concrete crack detection.

4 Conclusion

This study presents an intelligent system for automatic concrete crack detection by combining optimized pre-processing, ML classification, and real-time localization. The pre-processing stage consists of Gray scale conversion, CLAHE, noise removal and normalization. After that the PSNR and RMSE value achieved are 33.75 dB and 3.11 respectively. Classification accuracy improved progressively from SVM (99.27%), to SVM+RF+GB, which achieved the highest accuracy of 99.81%. Additionally, the YoLov7 model provided effective crack localization.

References

[1] Yu, Z.: YOLO V5s-based deep learning approach for concrete cracks detection. In: SHS Web of Conferences, vol. 144, p. 03015. EDP Sciences (2022)

[2] Nyathi, M.A., Bai, J., Wilson, I.D.: Deep learning for concrete crack detection and measurement. Metrology 4(1), 66–81 (2024)

[3] Wang, R., Chen, R., Yan, H., Guo, X.: Lightweight concrete crack recognition model based on improved MobileNetV3. Scientific Reports 15(1), 15704 (2025)

[4] Khan, M.A.M., Kee, S.H., Pathan, A.S.K., Nahid, A.A.: Image processing techniques for concrete crack detection: A scientometrics literature review. Remote Sensing 15(9), 2400 (2023)

[5] Hota, C.P.P.K., Sowjanya, V.V.N.L.S., Sree, G.R., Neeharika, K., Supritha, G.: Automatic detection and analysis of concrete cracks using YOLO. In: Proceedings of the 2024 4th International Conference on Intelligent Technologies (CONIT), Karnataka,India,pp.1–6 (2024). https://doi.org/10.1109/CONIT61985.2024.10627445.

[6] Harb, S., Achanccaray, P., Maboudi, M., Gerke, M.: Multi-temporal crack segmentation in concrete structures using deep learning approaches. arXiv preprint arXiv:2411.04620 (2024).

[7] Shibano, K., Morozova, N., Shimamoto, Y., Alver, N., Suzuki, T.: Improvement of crack detectivity for noisy concrete surface by machine learning methods and infrared images. Case Studies in Construction Materials 20, e02984 (2024).

[8] Choi, Y., Park, H.W., Mi, Y., Song, S.: Crack detection and analysis of concrete structures based on neural network and clustering. Sensors 24(6), 1725 (2024).

[9] Chen, S., Fan, G., Li, J., Hao, H.: Automatic complex concrete crack detection and quantification based on point clouds and deep learning. Engineering Structures 327, 119635 (2025).

[10] Shi, T., Luo, H.: Deep learning for automated detection and classification of crack severity level in concrete structures. Construction and Building Materials 472, 140793 (2025).

[11] https://www.kaggle.com/datasets/arunrk7/surface-crack-detection.