Indonesian Journal of Science and Technology

Journal homepage: http://ejournal.upi.edu/index.php/ijost/

Deep Learning Based Framework for Automated Road Crack Detection and Severity Assessment under Different Weather Conditions

Nguyen Nang An[1], Vinh V. Le[2], Nguyen Quang Lich[3], HongGiang Nguyen[3]*

[1]Hanoi Pedagogical University 2, PhuTho Province, Vietnam
[2]Posts and Telecommunications Institute of Technology, Hanoi City, Vietnam , PhuTho Province, Vietnam
[3]School of Engineering and Technology of Hue University, Hue City, Vietnam
*Correspondence: E-mail: giangnh@hueuni.edu.vn

ABSTRACT

Crack detection and assessment of roads and concrete are vital for infrastructure and safety management. This paper proposed an end-to-end deep learning based framework for automated crack detection, classification, and severity analysis under different surface conditions. The simulation implemented and tested four architectures, namely CNN, BiGRU, BiLSTM, and a hybrid BiGRU–BiLSTM–CNN model. Using a large dataset of 5,200 crack and non-crack photos collected from the METU campus and road photos taken using a smartphone, we trained and assessed our approach. Regression and classification metrics such as RMSE, MAE, R2, Cohen's Kappa, F1-score, and AUC were also used in the study to evaluate model performance. The CNN outperformed recurrent and hybrid models in capturing complex spatial crack information, according to the experimental data (R^2 = 0.92, F1-score = 0.97, AUC = 0.99). The suggested framework included geometric crack analysis and 3D heat map visualization in addition to probabilistic crack classification to measure relative depth and crack shape. The result offers an approach to real-time road condition evaluation using conventional cameras that is clear, dependable, and reasonably priced.

ARTICLE INFO

Article History:
Submitted/Received 03 Jan 2026
First Revised 15 Mar 2026
Accepted 27 Apr 2026
First Available Online 29 Apr 2026
Publication Date 30 Apr 2026

Keyword:
Crack geometry analysis,
Deep learning,
Dry and wet surface conditions,
Road crack detection.

1. INTRODUCTION

Road infrastructure degradation is the biggest issue facing modern transport systems, as surface cracking is the most visible indicator of structural decline [1-3]. If unchecked, cracks can propagate to the underlying road layers, resulting in rapid deterioration, reduced vehicle control, reduced ride quality, and increased maintenance/repair costs [4-5]. Existing techniques, including visual inspection, ground-penetrating radar, or laser scans, also face problems of operational costs, manual involvement, and vulnerability to lighting, weather, and human inconsistencies [6-8]. Likewise, conventional machine learning approaches, such as support vectors, random forests, K-nearest neighbors, and decision trees, rely on manual feature extraction and cannot self-learn high-dimensional crack representations [9-10]. New developments in artificial intelligence (AI), especially deep learning, have significantly increased road surface monitoring. CNNs allow for spatial feature learning by autonomously learning crack attributes like width, complexity of edges, and branching. Recurrent models, such as BiLSTMs, can model sequential and temporal crack characteristics but generally demand significant computational overhead. In contrast, BiGRUs offer similar temporal modeling performance with much less computational and memory requirements, making them better suited for real-time applications. Additionally, the hybrid architectures that combine CNNs with BiGRU and BiLSTM models are more robust to environmental noise, surface roughness, and nonlinear crack propagation [11-13]. Compared with classical machine learning and pure deep learning approaches, experimental results indicate that such hybrid approaches attain better accuracy, flexibility, and automation [14-16].

Recent studies have demonstrated the outstanding potential of artificial intelligence (AI), particularly deep learning, in automating crack detection and road condition monitoring with high accuracy and real-time deployment capabilities. The advanced performance of the DCU Former model in surface crack segmentation, while also highlighting challenges such as data imbalance and practical implementation [17]. A real-time crack detection system based on MobileNetV2 and XGBoost, achieving 99% accuracy and 42 fps speed on an embedded platform [18]. Surveys by previous study, all affirm the superiority of deep learning methods over traditional image processing, while highlighting remaining challenges such as generality, real-time deployment, and integration with smart sensors [19-22].

While prior work has broadly leveraged deep learning for crack detection, nearly all have applied CNN-based models and emphasized classification accuracy, with few comparisons of different architectures or evaluations in diverse environments. To fill this gap, this research evaluates CNN, BiGRU, BiLSTM, and a hybrid model for wet and dry road crack detection by combining quantitative metrics with geometric crack analysis.

Based on this overview, the study proposes an integrated spatiotemporal deep learning framework that combines quantitative geometric assessments of cracks under different environmental conditions (dry and wet) with synchronous heat mapping and contour analysis. The results show that the proposed method allows for early detection of microcracks and accurate prediction of propagation trends, thereby supporting proactive road maintenance, optimizing lifecycle costs, and improving traffic safety.

2. METHODS

To accurately detect and classify crack patterns, four deep learning architectures such as CNN, BiGRU, BiLSTM, and a hybrid BiGRU-BiLSTM-CNN model were implemented. Each model capitalizes on a different feature extraction capability to boost prediction prowess. At the same time, the experiments were implemented in Python 3.12.4 using several scientific and

DOI: https://doi.org/10.17509/ijost.v11i1.95376
p- ISSN 2528-1410 e- ISSN 2527-8045

deep learning libraries. OpenCV and PIL, NumPy, Matplotlib, TensorFlow/Keras, and Scikit-learn. All experiments were conducted on an ASUS laptop (8th Gen Intel Core i7, 32 GB RAM.

2.1. Convolutional Neural Network (CNN) Model

CNNs are good at capturing local spatial features like texture and edges, which makes them well suited for crack pattern recognition [23-24]. The CNN architecture for the study in **Figure 1** includes three hierarchical convolutional blocks:

Conv2D → ReLU → MaxPooling (repeated three times)
Flatten
Dense(64) → ReLU → Dropout(0.5)
Dense(1) → Sigmoid

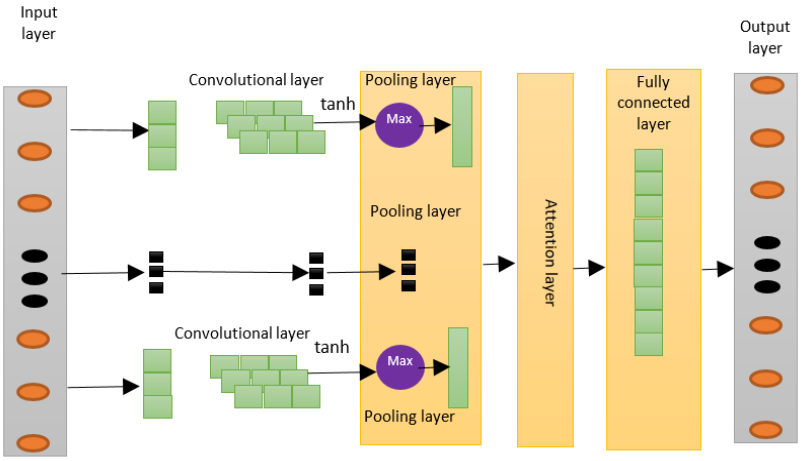


Figure 1. The CNN structure.

2.1. Bidirectional Gated Recurrent Unit (BiGRU) Model

BiGRUs learn temporal relationships in sequences from images with less computation than LSTMs. Their bidirectional nature enables processing feature patterns in both directions. The sructural model in **Figure 2** is decribed as below:

BiGRU(128) → BatchNorm → Dropout(0.5)
BiGRU(64) → BatchNorm → Dropout(0.4)
Dense(128, ReLU) → Dropout(0.3)
Dense(64, ReLU) → Dropout(0.3)
Dense(1, Sigmoid)

DOI: https://doi.org/10.17509/ijost.v11i1.95376
p- ISSN 2528-1410 e- ISSN 2527-8045

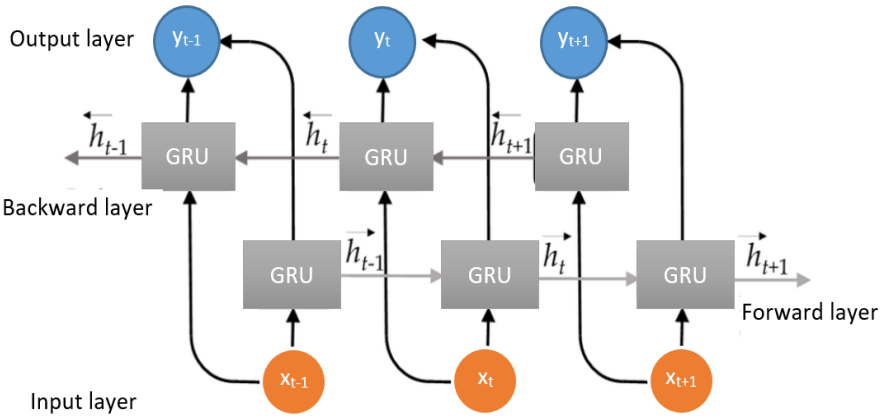


Figure 2. BiGRU structure.

2.3. Bidirectional Long Short-Term Memory (BiLSTM) Model

BiLSTMs excel at learning long-range dependencies, making them robust in modeling crack geometry trends and progression [25-26]. The architecture of the model in **Figure 3** show as:

BiLSTM(128) → BatchNorm → Dropout(0.4)
BiLSTM(64) → BatchNorm → Dropout(0.4)
Dense(128, ReLU) → Dropout(0.3)
Dense(64, ReLU) → Dropout(0.3)
Dense(1, Sigmoid)

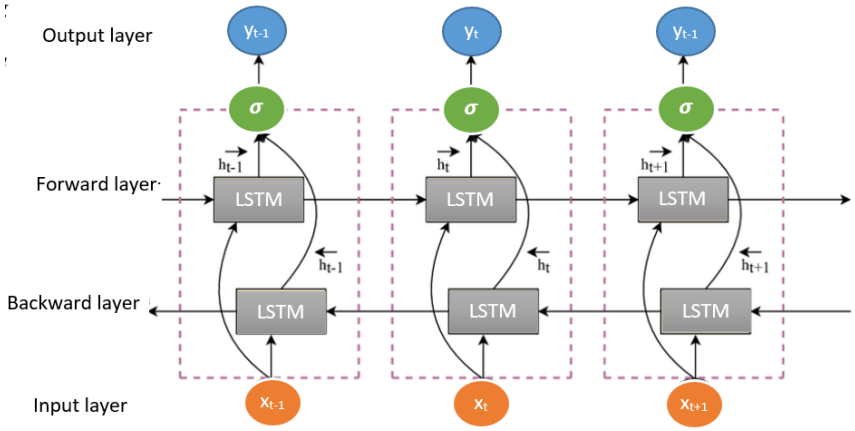


Figure 3. The BiLSTM structure.

2.4. Hybrid BiGRU–BiLSTM–CNN Model

A hybrid BiGRU–BiLSTM–CNN architecture is designed to combine temporal sequence learning with spatial crack feature representation. First, BiGRU and BiLSTM layers perform sequential learning to model crack propagation dynamics and long–short term dependencies. Next, a feature expansion stage transforms learned representations into structured feature maps. Finally, CNN layers extract spatial textures and morphological features of cracks for robust classification. The novelty of this approach lies in integrating temporal sequence learning with spatial convolutional analysis, enabling the model to simultaneously capture

DOI: https://doi.org/10.17509/ijost.v11i1.95376
p- ISSN 2528-1410 e- ISSN 2527-8045

crack geometry and propagation behavior, thereby improving prediction accuracy and feature interpretability. The model configuration for the simulation in **Figure 4** consists of:

Sequence learning
BiGRU(128, dropout=0.2) → BatchNorm
BiLSTM(64, dropout=0.2) → BatchNorm
Feature expansion
Dense(256, ReLU) → Dropout(0.3)
Dense(64, ReLU) → Reshape(8×8×1) → BatchNorm
Spatial analysis
Conv2D(32) → MaxPooling
Conv2D(64) → MaxPooling
Flatten
Final classification
Dense(128, ReLU) → Dropout(0.3)
Dense(64, ReLU) → Dropout(0.3)
Dense(1, Tanh)

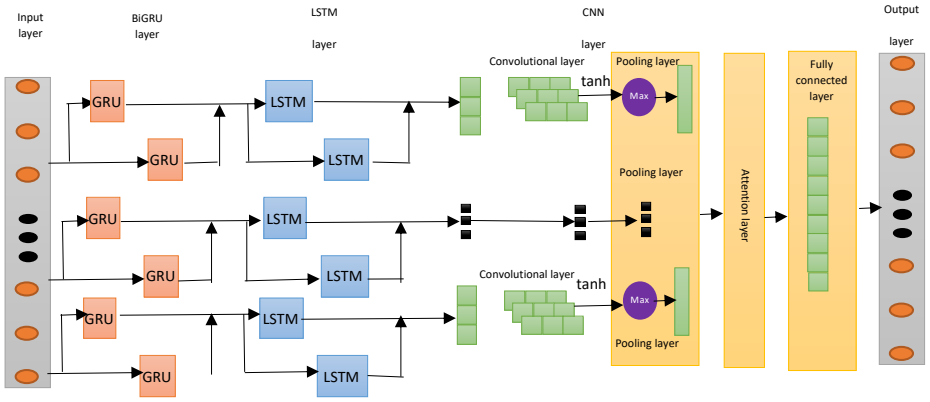


Figure 4. Hybrid structure.

Table 1. Models' configuration indiates the network architectures and training configurations of our four deep learning models, including CNN, BiGRU, BiLSTM, and our hybrid BiGRU–BiLSTM–CNN model. The table summarizes our layer structures, key parameters, and typical training settings, giving a nice overview of the processing framework we implemented for road crack detection.

Table 1. Models' configuration.

Model	Layer / Stage	Configuration / Parameters	Function
CNN	Input	Image size: 256 × 256 × 3 (RGB)	Road crack image input
	Conv2D	32 filters, kernel 3×3, ReLU	Low-level spatial feature extraction
	MaxPooling	2 × 2	Downsampling
	Conv2D	64 filters, kernel 3×3, ReLU	Crack texture extraction
	MaxPooling	2 × 2	Feature reduction
	Conv2D	128 filters, kernel 3×3, ReLU	High-level spatial features
	MaxPooling	2 × 2	Dimensionality reduction
	Flatten	—	Convert feature map to vector

DOI: https://doi.org/10.17509/ijost.v11i1.95376
p- ISSN 2528-1410 e- ISSN 2527-8045

Table 1 (Continue). Models' configuration.

Model	Layer / Stage	Configuration / Parameters	Function
	Dense	128 units, ReLU	Feature representation
	Dense	64 units, ReLU	Feature compression
	Output	Dense (1), Sigmoid	Binary crack classification
BiGRU	BiGRU	128 units	Temporal sequence learning
	BatchNorm	Default	Training stabilization
	Dropout	0.4	Regularization
	BiGRU	64 units	Deeper sequential learning
	BatchNorm	Default	Feature normalization
	Dropout	0.4	Overfitting reduction
	Dense	128 units, ReLU	Feature transformation
	Dropout	0.3	Regularization
	Dense	64 units, ReLU	Feature compression
	Output	Dense (1), Sigmoid	Crack prediction
BiLSTM	BiLSTM	128 units	Temporal dependency learning
	BatchNorm	Default	Stabilize training
	Dropout	0.4	Regularization
	BiLSTM	64 units	Deep sequential modeling
	BatchNorm	Default	Feature normalization
	Dropout	0.4	Overfitting reduction
	Dense	128 units, ReLU	Feature representation
	Dropout	0.3	Regularization
	Dense	64 units, ReLU	Feature compression
	Output	Dense (1), Sigmoid	Binary crack classification
Hybrid BiGRU–BiLSTM–CNN	BiGRU	128 units, dropout 0.2	Sequential learning
	BatchNorm	Default	Stabilize training
	BiLSTM	64 units, dropout 0.2	Long–short term dependency learning
	BatchNorm	Default	Feature normalization
	Dense	256 units, ReLU	Feature expansion
	Dropout	0.3	Regularization
	Dense	64 units, ReLU	Feature transformation
	Reshape	8 × 8 × 1	Convert sequence features to spatial map
	Conv2D	32 filters, kernel 3×3	Spatial texture extraction
	MaxPooling	2 × 2	Spatial reduction
	Conv2D	64 filters, kernel 3×3	Deep spatial feature learning
	MaxPooling	2 × 2	Feature abstraction
	Flatten	—	Convert feature map to vector
	Dense	128 units, ReLU	High-level feature representation
	Dropout	0.3	Overfitting reduction
	Dense	64 units, ReLU	Feature compression
	Output	Dense (1), Tanh	Final crack prediction
Common Training Settings (All Models)	Optimizer	Adam	Gradient optimization
	Learning rate	0.001	Model convergence control
	Batch size	32	Training efficiency
	Epochs	20	Model training iterations
	Loss function	Binary cross-entropy	Classification loss
	Evaluation metrics	RMSE, MAE, R^2, F1-score, AUC, Cohen's Kappa	Performance evaluation

DOI: https://doi.org/10.17509/ijost.v11i1.95376
p- ISSN 2528-1410 e- ISSN 2527-8045

2.5. Accuracy Metric

Predicted accuracy indicators work by tracking and comparing actual and forecast data. These metrics evaluate the predicted quantity in the research, namely RMSE, MAE, R-squared (R^2), Kappa, F1 score, AUC. The indicators are presented in Eqs. (1-6).

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(x_t - x_t')^2}{n}} \quad (1)$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|x_t' - x_t| \quad (2)$$

$$R\_squared = 1 - \frac{\sum_{t=1}^{n}(x_t - x_t')^2}{\sum_{t=1}^{n}\left(x_t - \frac{1}{n}\sum_{t=1}^{n} x_t\right)^2} \quad (3)$$

$$\text{Kappa} = \frac{p_0 - p_e}{1 - p_e} \quad (4)$$

$$\text{F1 score} = \frac{Precision\ x\ Recall}{Precision + Recall} \quad (5)$$

$$\text{AUC} = \int_0^1 TPR(FPR)d(FPR) \quad (6)$$

where x_t, x_t' name the predicted and actual values in the period time t, and n is the number of the observed data in the testing stage, $\bar{x}$: $\overline{x'}$ describe the mean of the predicted and mean of the actual values; p_0, p_e are observed agreement, and expected agreement by chance; $Precision = \frac{TP}{TP+FP}$' and $Recall = \frac{TP}{TP+FN}$' TP (True Positives) denotes the number of correctly predicted positive instances, FP (False Positives) refers to the number of instances incorrectly predicted as positive, FN (False Negatives) represents the number of instances incorrectly predicted as negative, TPR (True Positive Rate) also known Recall, measures the proportion of actual positives correctly identified, FPR (False Positive Rate) measures the proportion of actual negatives that are incorrectly classified as positive, and TN (True Negatives) denotes the number of correctly predicted negative instances.

2.6. Data Collection

The dataset used in this study contains raod surface crack images, collected at different buildings across the Middle East Technical University (METU) campus in Ankara, Turkey (dataset can download at *https://data.mendeley.com/datasets/5y9wdsg2zt/1*). As for image classification, the data is divided into two classes: crack (positive) and non-crack (negative). Each class has 20,000 images, for 40,000. All images are 227 × 227 pixels in three RGB channels. Also, 10 images were taken on a Samsung A12 phone in wet and dry road surface conditions to detect the model in random roads in Hue City Vietnam.

Due to hardware limitations, including a Core i7 CPU and 32 GB of RAM, only a subset of 5,200 images of the total 40,010 images was selected for simulation and model training. Specifically, for each class, 2,000 images were used for training, 300 images for validation, and 300 images for testing, corresponding to 70%, 15%, and 15% of the dataset, respectively. All images were loaded and preprocessed prior to model construction and performance evaluation.

Figure 5 introduces an end-to-end structural simulation pipeline for automated crack detection and analysis. These preprocessed images are assessed with several deep learning models like CNN, BiGRU, BiLSTM, and hybrids. Model performance is used to compare and

DOI: https://doi.org/10.17509/ijost.v11i1.95376
p- ISSN 2528-1410 e- ISSN 2527-8045

pick the best predictor. The chosen model does probabilistic crack classification, then geometric crack analysis, visualization by heat maps, and statistical reporting to aid in overall road condition assessment.

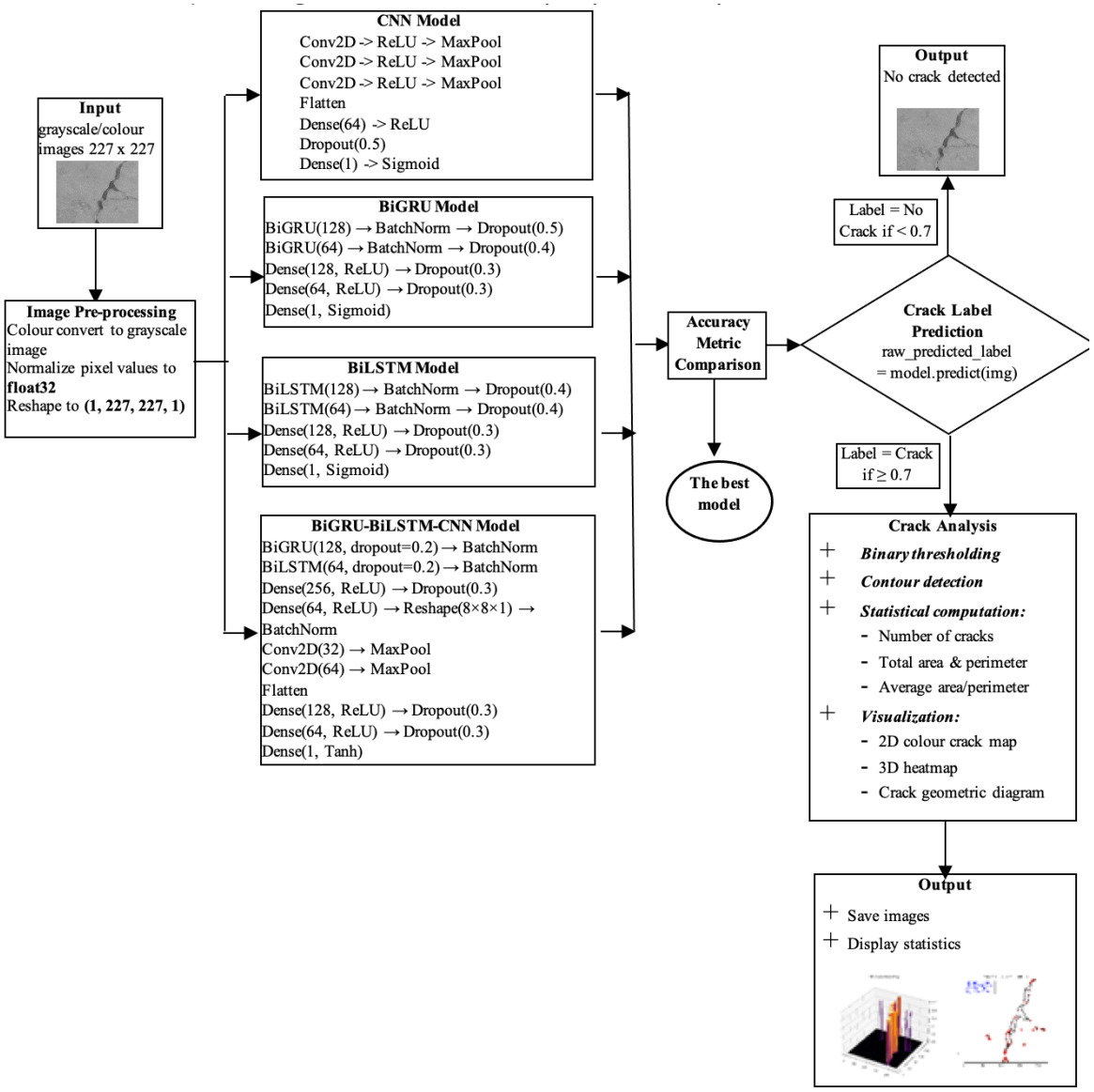


Figure 5. The general simulation of the study.

3. RESULTS AND DISCUSSION

This describes our study on the performance of three deep learning standalone models: CNN, BiGRU, and BiLSTM, along with a hybrid architecture that combines them to simulate and evaluate road or road crack severity under dry and wet surface conditions. **Figure 6** illustrates the training and validation loss convergence behavior of the four models over 20 epochs. The CNN model converges fastest and most stably, reaching the lowest loss values during training. Instead, the BiLSTM and hybrid models exhibit moderate convergence, with the BiGRU model stabilizing more slowly.

As shown by the quantitative performance metrics summarized in **Table 2**, the CNN model achieves the best overall performance among the approaches. Notably, it also logs the minimal error (RMSE = 0.14, MAE = 0.03), maximal R^2 (0.92), and exceptional classification results, with Cohen's Kappa (0.94), F1-score (0.97), and AUC (0.99). Subsequently, the best is

DOI: https://doi.org/10.17509/ijost.v11i1.95376
p- ISSN 2528-1410 e- ISSN 2527-8045

the BiLSTM, which also demonstrates robust predictive power (RMSE = 0.19, MAE = 0.06, R^2 = 0.85) and classification accuracy (Kappa = 0.90, F1-score = 0.95, AUC = 0.99). While the hybrid model slightly bettered BiGRU on most metrics, both methods lagged behind CNN and BiLSTM, indicating a deficiency in capturing intricate spatial crack characteristics.

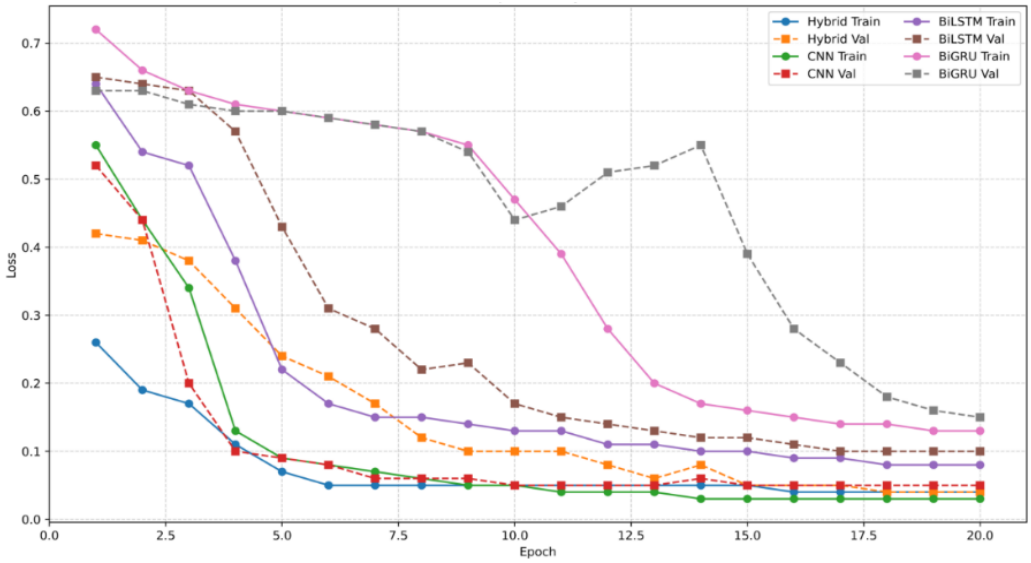


Figure 6. Training and validation loss curves of the models.

Table 2. Performance metrics of the models for road crack detection.

Model	RMSE	MAE	R^2	Kappa	F1 score	AUC
CNN	0.14	0.03	0.92	0.94	0.97	0.99
BiGRU	0.26	0.10	0.72	0.83	0.92	0.98
BiLSTM	0.19	0.06	0.85	0.90	0.95	0.99
Hybrid	0.20	0.06	0.84	0.89	0.95	0.99

Figures 7 and **8** presents a qualitative comparison of crack detection results generated by CNN, BiLSTM, BiGRU, and hybrid models. For each method, the original image and various preprocessing results, including grayscaling, binary thresholding, Gaussian blur, truncation, Canny, Sobel X, and Laplacian, are shown. The CNN model generates the most realistic and uniform crack depictions, in line with its top R^2 score. The BiLSTM and hybrids appear visually similar; however, the BiGRU still exhibits weaker crack delineation. These visual observations are further supported by the quantitative evaluation metrics.

To identify crack occurrence during simulation, the authors created two complementary prediction functions by combining deep learning with traditional image processing. The initial function performs crack classification using a pretrained deep learning model. The input images are normalized to the network architecture, and the model returns a continuous probability of crack. A threshold of 0.7 is then used to convert this probability to a binary label (Crack/No Crack), with confidence levels being considered based on the magnitude of the probability to enable confident decision-making. The second function quantitatively analyzes crack morphology based on binary images by extracting contours and calculating the number of cracks, total crack area, total perimeter, and related averages. This two-pronged approach facilitates not only identifying cracks but also quantifying their severity in a quantitative manner.

DOI: https://doi.org/10.17509/ijost.v11i1.95376
p- ISSN 2528-1410 e- ISSN 2527-8045

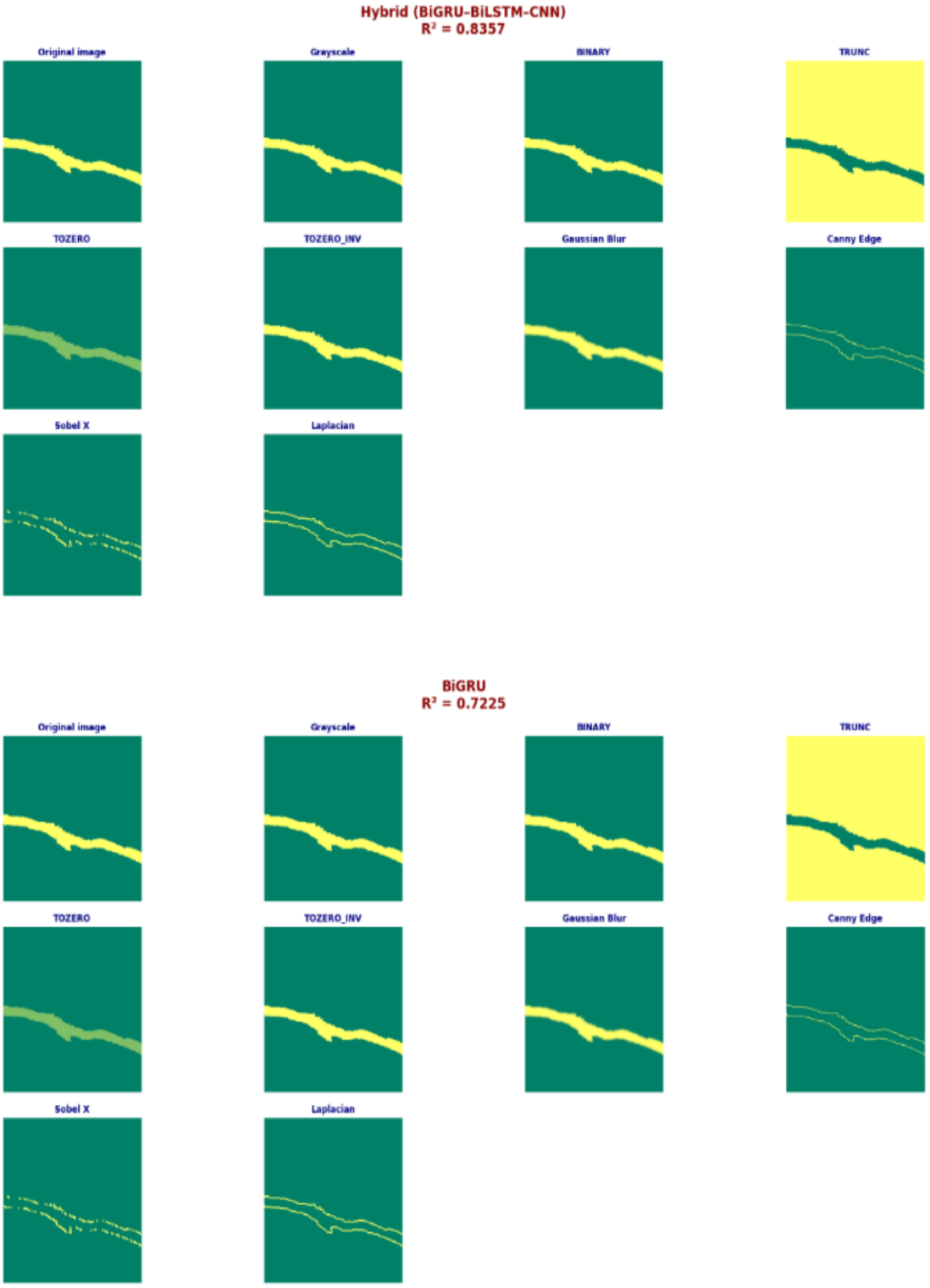


Figure 7. Visual comparison of crack detection results using hybrid, and BiGRU models.

Model validation with randomly selected images from the test set demonstrates that all four models distinguish between cracked and non-cracked conditions in photos captured of building cracks on the METU campus. The images in **Table 3** demonstrate the visualization results of crack detection and geometric analysis for a representative positive case and a negative condition. For confirmation, randomly selected test images were examined by the trained models. In the positive case, the model successfully detects a crack with a prediction confidence of 0.7928. The identified crack area is then isolated via crack coloring, in which

DOI: https://doi.org/10.17509/ijost.v11i1.95376
p- ISSN 2528-1410 e- ISSN 2527-8045

pixels labeled as cracks are colored. Based on this segmentation mask, geometric properties are computed using contour extraction, including a total crack area of 197.0 pixel², total perimeter of 723.6 pixels, average area of 90.8 pixel², and average perimeter of 32.9 pixels. The 3D Crack Heat Map is generated by converting the 2D crack probability map into a height representation, where pixel intensity values correspond to normalized crack activation levels (0–1). This visualization highlights the variations in crack severity and distribution over the surface. On the other hand, negative samples do not exhibit crack segmentation or geometry, validating the model's identification of crack-free surfaces.

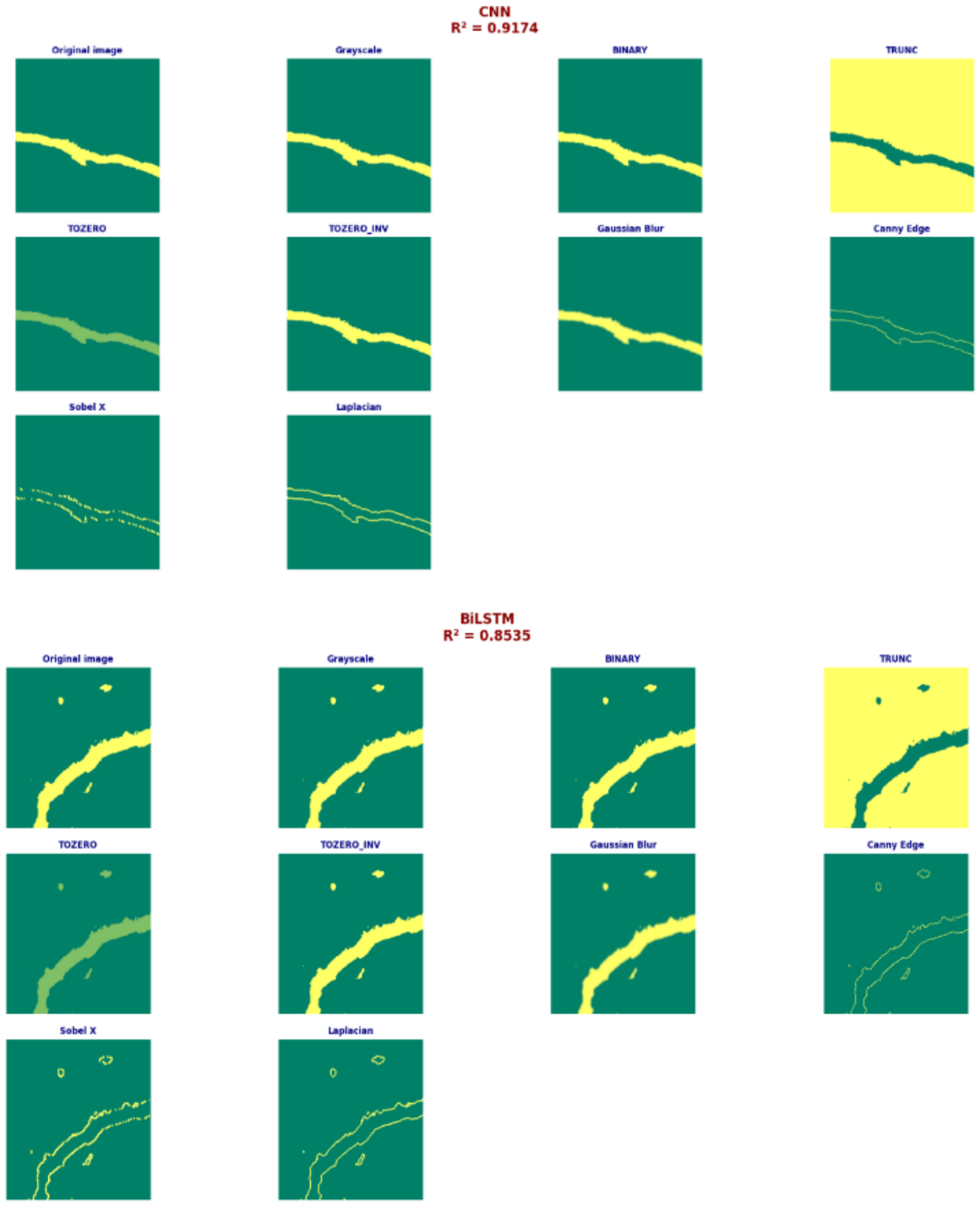
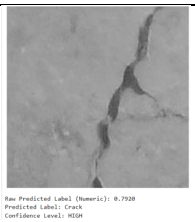
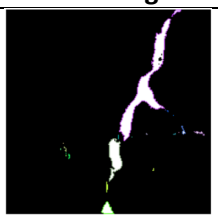
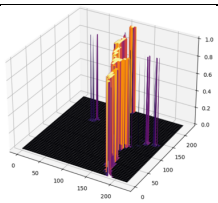
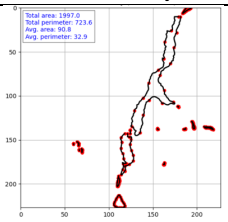
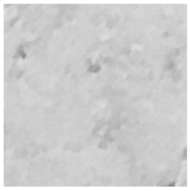


Figure 8. Visual comparison of crack detection results using CNN, and BiLSTM models.

DOI: https://doi.org/10.17509/ijost.v11i1.95376
p- ISSN 2528-1410 e- ISSN 2527-8045

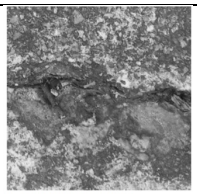
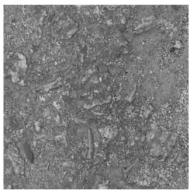
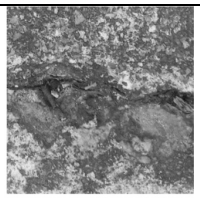
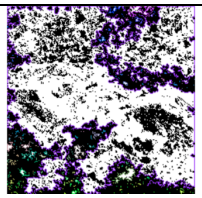
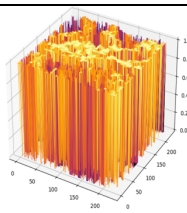
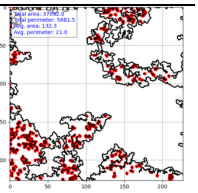
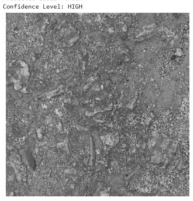
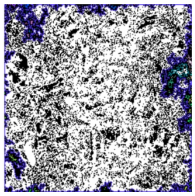
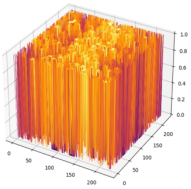
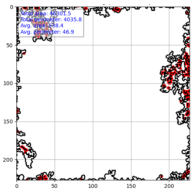
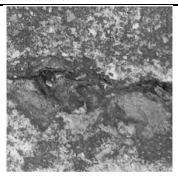
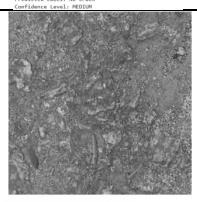
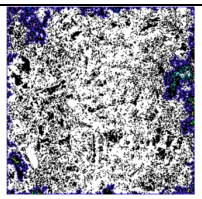
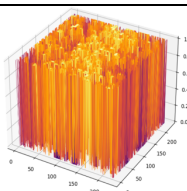
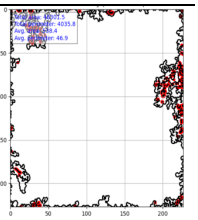
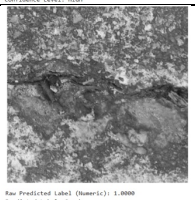
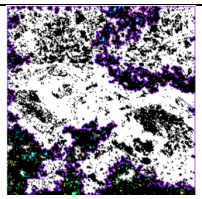
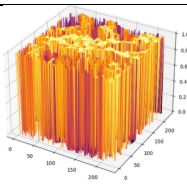
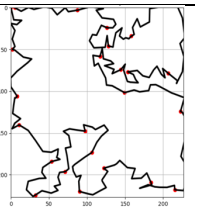
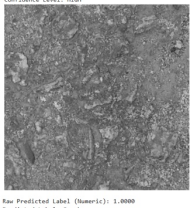
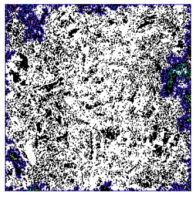
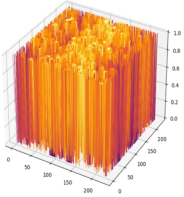
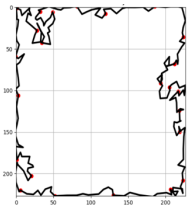
For road crack evaluation, images were captured using a Samsung A12 smartphone without any brightness or resolution modifications to preserve real field conditions. The dataset comprises images under both dry and wet surface conditions, as summarized in **Tables 4** and **5**. Under dry conditions, crack features stand out with relatively low background noise; therefore, the models can easily locate crack patterns. The CNN and BiLSTM models detect positive crack samples with confidence values around 0.79. After detection, a crack segmentation map is generated by highlighting pixels classified as crack regions. From this segmentation mask, geometric features are extracted using contour detection, including crack area and perimeter, which show values similar to those obtained from building crack experiments. Negative samples return almost zero geometric markers, verifying that no cracks exist. The 3D Crack Heat Map is obtained by transforming the 2D crack probability or segmentation map into a three-dimensional surface representation. Specifically, the pixel intensity values of the crack activation map are normalized within the range of 0 and 1, and these values are then interpreted as the height (z-axis) in a 3D visualization. Consequently, areas of pronounced crack response emerge as raised peaks in the 3D plot, representing the spatial distribution and intensity of cracking over the surface. Under wet conditions (see **Table 5**), crack detection becomes more challenging due to reduced contrast, water reflections, and higher background noise. Nevertheless, the CNN, BiLSTM, and BiGRU models still detect cracks with confidence levels within the range of 0.75 and 0.80. The corresponding geometric analysis indicates total crack areas ranging from 190–200 pixel² and perimeters exceeding 700 pixels, whereas negative samples continue to exhibit minimal crack traces.

Table 3. Visualization results of crack detection and geometric analysis under different conditions.

Situation	Orginal image	Crack Map with Heat Coloring	3D Crack Heat Map	Crack Geometry (All Contours)
Positive				
Negative		NA	NA	NA

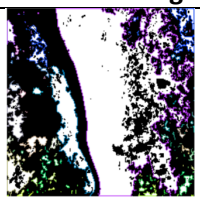
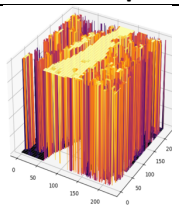
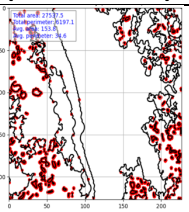
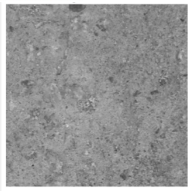
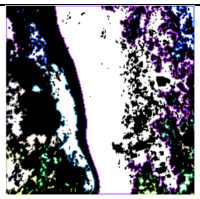
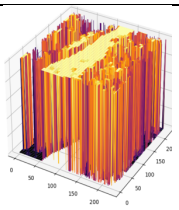
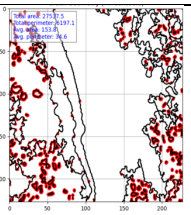
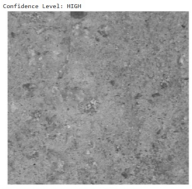
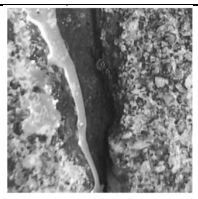
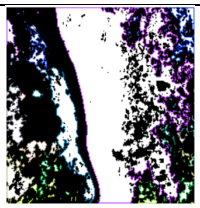
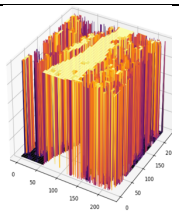
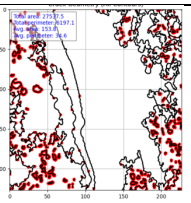
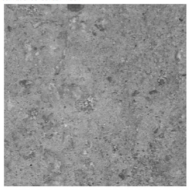
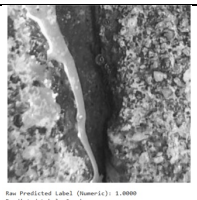
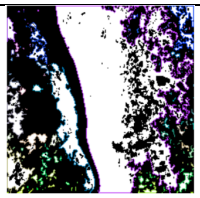
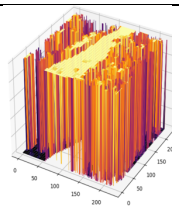
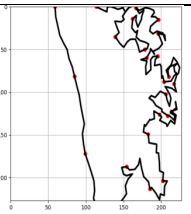
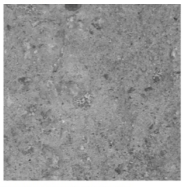
DOI: https://doi.org/10.17509/ijost.v11i1.95376
p- ISSN 2528-1410 e- ISSN 2527-8045

Table 4. Visualization results of crack detection and geometric analysis under dry condition.

Model	Situation	Orginal image	Crack Map with Heat Coloring	3D Crack Heat Map	Crack Geometry (All Contours)
Hybrid	Positive		NA	NA	NA
	Negative		NA	NA	NA
BiLSTM	Positive				
	Negative				
BiGRU	Positive		NA	NA	NA
	Negative				
CNN	Positive				
	Negative				

DOI: https://doi.org/10.17509/ijost.v11i1.95376
p- ISSN 2528-1410 e- ISSN 2527-8045

Table 5. Visualization results of crack detection and geometric analysis under wet condition.

Model	Situation	Orginal image	Crack Map with Heat Coloring	3D Crack Heat Map	Crack Geometry (All Contours)
Hybrid	Positive				
	Negative		NA	NA	NA
BiLSTM	Positive				
	Negative		NA	NA	NA
BiGRU	Positive				
	Negative		NA	NA	NA
CNN	Positive				
	Negative		NA	NA	NA

DOI: https://doi.org/10.17509/ijost.v11i1.95376
p- ISSN 2528-1410 e- ISSN 2527-8045

Several prior studies have investigated crack detection using deep learning techniques; however, most of them primarily focus on convolutional neural network (CNN) architectures for pavement and structural crack identification. For instance, Fan et al. [27] proposed a CNN-based structured prediction framework for automated pavement crack detection, demonstrating that CNN models can effectively learn spatial crack patterns and achieve classification accuracies exceeding 90% (F1-score = 0.92). Similarly, Ali et al. [28] evaluated several deep CNN architectures, including VGG-16 and ResNet-50, for crack detection in concrete structures, reporting accuracies between 92% and 96%.

Another related study by Matarneh et al. [29] developed a transfer learning approach that was applied with multiple pre-trained CNN architectures to classify 2,139 pavement crack images into three crack categories. Of the tested models, DenseNet201 had the highest baseline accuracy (94.12%). Once Grey Wolf Optimizer (GWO) was implemented, the accuracy increased to 98.73%, showcasing the power of optimization. In contrast to these earlier efforts, the present study provides a more comprehensive evaluation framework by examining three standalone deep learning models, namely CNN, BiGRU, and BiLSTM, together with a hybrid architecture. Unlike many previous studies that focus mainly on classification accuracy, this research incorporates multiple quantitative performance metrics, including RMSE, MAE, R^2, F1-score, AUC, and Cohen's Kappa, enabling a more robust and multidimensional assessment of model performance. The experimental results indicate that the CNN model achieves the best overall performance (RMSE = 0.14, MAE = 0.03, R^2 = 0.92, F1-score = 0.97), highlighting its strong predictive capability in capturing spatial crack features.

Furthermore, this study extends previous work by integrating geometric crack analysis and visualization, including crack area, perimeter, contour extraction, and 3D crack heat maps derived from normalized probability maps. The evaluation also considers crack detection under both dry and wet road surface conditions, which better reflects realistic monitoring scenarios. Experimental validation displays that all four models produce consistent predictions for images collected on the METU campus, whereas smartphone images captured using a Samsung A12 demonstrate performance variations due to environmental conditions. Under dry conditions, all models illustrate high accuracy in detecting non-cracked surfaces, while CNN and BiLSTM outperform in crack detection. Under wet conditions, all models also keep comparable prediction performance, demonstrating great robustness to noise and illumination changes.

Overall, the findings confirm that the CNN model outperforms recurrent-based approaches in capturing complex spatial crack characteristics (R^2 = 0.92). Moreover, validation with raw smartphone images highlights the practical feasibility, low deployment cost, and significant potential of the proposed framework for real-time road condition monitoring and intelligent maintenance decision support.

4. CONCLUSION

This paper introduced a novel failure detection and ranking framework that combines deep learning with geometric and depth-based crack analysis. Leveraging a large-scale dataset of 5,200 in 40,010 images sampled from buildings on METU campus and further smartphone images captured in dry and wet conditions, the proposed approach proves to be robust and applicable to real-world scenarios. Among the models tested, CNN consistently performed best across regression and classification metrics (R^2 = 0.92, F1-score = 0.97, AUC = 0.99), emphasizing its effectiveness in capturing complex spatial crack features. The combination of

DOI: https://doi.org/10.17509/ijost.v11i1.95376
p- ISSN 2528-1410 e- ISSN 2527-8045

probabilistic classification, quantitative geometric indices, and 3D depth visualization enables accurate crack identification while providing a reliable assessment of damage severity.

Even with promising results, there are still some limitations. First, the data was gathered in a relatively narrow context and not necessarily generalizable to other road types or environments. Second, the computer setup, with its restricted scope, could not digest all of the data.

Future work will expand the dataset to different road materials and locations, include real-world depth sensors for actual crack depth estimation, and attempt advanced architectures like attention or transformers. In addition, real-time deployment on mobile or edge devices will be explored to enable large-scale, low-cost infrastructure monitoring.

5. AUTHORS' NOTE

The authors declare that there is no conflict of interest regarding the publication of this article. The authors confirmed that the paper was free of plagiarism.

6. REFERENCES

[1] Milling, A., Martin, H., and Mwasha, A. (2023). Design, construction, and in-service causes of premature pavement deterioration: a fuzzy Delphi application. *Journal of Transportation Engineering, Part B: Pavements*, *149*(1), 05022004.

[2] Rincon, L. F., Moscoso, Y. M., Hamami, A. E. A., Matos, J. C., and Bastidas-Arteaga, E. (2024). Degradation models and maintenance strategies for reinforced concrete structures in coastal environments under climate change: a review. *Buildings*, *14*(3), 562.

[3] Septiyani, Y. N. (2024). The impact of load traffic of road deterioration in urban areas: case study Jalan KH Abdul Halim Majalengka. LEADER: *Civil Engineering and Architecture Journal*, *2*(4), 911-919.

[4] Bhosale, T., Attar, A., Warang, P., and Patil, R. (2022). Object detection for autonomous guided vehicle*. ASEAN Journal of Science and Engineering*, *2*(3), 209-216.

[5] Pohan, M. A. R., Utama, J., and Herdiana, B. (2024). Novel motion planning strategy with fuzzy logic for improving safety in autonomous vehicles in response to risky road user behaviors. *ASEAN Journal of Science and Engineering*, *4*(3), 471-484.

[6] Maske, M. (2022). Review of applications of ground penetrating radar as an NDT tool. *ASEAN Journal of Science and Engineering*, *2*(2), 115-128.

[7] Merkle, D., Frey, C., and Reiterer, A. (2021). Fusion of ground penetrating radar and laser scanning for infrastructure mapping. *Journal of Applied Geodesy*, *15*(1), 31-45.

[8] Rangole, A., De, S., Kuchekar, N., and Raj, A. B. (2024). A comprehensive review of ground penetrating radar: Techniques, applications and future directions. *International Journal of Engineering Research and Reviews*, 12(3), 30-53.

[9] Shayea, G. G., Zabil, M. H. M., Habeeb, M. A., Khaleel, Y. L., and Albahri, A. (2025). Strategies for protection against adversarial attacks in AI models: An in-depth review. *Journal of Intelligent Systems*, *34*(1), 20240277.

[10] Yang, Y., Pan, Z., Sun, J., Welch, J., and Klionsky, D. J. (2024). Autophagy and machine learning: Unanswered questions. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, *1870*(6), 167263.

[11] Chen, Y., Kang, J., Feng, L., Yuan, L., Liang, J., Zhao, Z., and Wu, B. (2024). Deep learning-based frequency-multiplexing composite-fringe projection profilometry technique for one-shot 3D shape measurement. *Measurement*, *233*, 114640.

DOI: https://doi.org/10.17509/ijost.v11i1.95376
p- ISSN 2528-1410 e- ISSN 2527-8045

[12] Hadj-Attou, A., Kabir, Y., and Ykhlef, F. (2023). Hybrid deep learning models for road surface condition monitoring. *Measurement*, *220*, 113267.
[13] Huang, Z., Xu, G., Zhang, X., Zang, B., and Yu, H. (2025). Three-dimensional ground-penetrating radar-based feature point tensor voting for semi-rigid base asphalt pavement crack detection. *Developments in the Built Environment*, *21*, 100591.
[14] Afzal, A. (2024). A hybrid approach combining computer vision and machine learning for enhanced image analysis. *Academia Nexus Journal*, *3*(3), 1-15.
[15] Choudhary, G., and Sethi, D. (2023). From conventional approach to machine learning and deep learning approach: an experimental and comprehensive review of image fusion techniques. *Archives of Computational Methods in Engineering*, *30*(2), 1267-1304.
[16] Sajid, M., Malik, K. R., Almogren, A., Malik, T. S., Khan, A. H., Tanveer, J., and Rehman, A. U. (2024). Enhancing intrusion detection: a hybrid machine and deep learning approach. *Journal of Cloud Computing*, *13*(1), 123.
[17] Zhuang, H., Cheng, Y., Zhou, M., and Yang, Z. (2025). Deep learning for surface crack detection in civil engineering: A comprehensive review. *Measurement*, 116908.
[18] Thohari, A. N. A., Karima, A., Santoso, K., and Rahmawati, R. (2024). Crack detection in building through deep learning feature extraction and machine learning approch. *Journal of Applied Informatics and Computing*, *8*(1), 1-6.
[19] Khan, S., Jan, A., and Seo, S. (2023). Structural crack detection using deep learning: an in-depth review. Korean *Journal of Remote Sensing*, *39*(4), 371-393.
[20] El-Din Hemdan, E., and Al-Atroush, M. (2025). A review study of intelligent road crack detection: Algorithms and systems. *International Journal of Pavement Research and Technology*, *2025*, 1-31.
[21] Manoni, L., Orcioni, S., and Conti, M. (2024). Recent advancements in deep learning techniques for road condition monitoring: *A comprehensive review. IEEE Access*, *12*, 154271-154293.
[22] Gupta, P., and Dixit, M. (2022). Image-based crack detection approaches: a comprehensive survey. Multimedia Tools and Applications, *81*(28), 40181-40229.
[23] Birgani, S. A., Zadeh, S. S., Davari, D. D., and Ostovar, A. (2024). Deep Learning Applications for Analysing Concrete Surface Cracks. *International Journal of Applied Data Science in Engineering and Health*, *1*(2), 69-84.
[24] Rashid, T., Mokji, M. M., and Rasheed, M. (2025). Cracked concrete surface classification in low-resolution images using a convolutional neural network. *Journal of Optics*, *54*(5), 3671-3683.
[25] Vargas, I. G. (2024). Deep Learning Approaches for Defect Segmentation on Composite Materials using Infrared Thermography. https://repositorio.ufu.br/handle/123456789/44621
[26] Yang, S., Shi, H., Yin, J., Xu, X., Yao, J., and Liu, S. (2025). PLI and CNN-BiLSTM: An enhanced data augmentation and deep learning approach for defect recognition in spiral welded pipe based on ultrasonic guided waves. *IEEE Sensors Journal*, *25*(19), 35916-35929.
[27] Fan, Z., Y. Wu, J. Lu, and W. Li. (2018). Automatic pavement crack detection based on structured prediction with the convolutional neural network. *arXiv preprint arXiv*:1802.02208. DOI: https://doi.org/10.48550/arXiv.1802.02208.
[28] Ali, L., et al. (2021). Performance evaluation of deep CNN-based crack detection and localization techniques for concrete structures. *Sensors.* 21(5): p. 1688. DOI: https://doi.org/10.3390/s21051688.

DOI: https://doi.org/10.17509/ijost.v11i1.95376
p- ISSN 2528-1410 e- ISSN 2527-8045

[29] Matarneh, S., et al. (2024). Evaluation and optimisation of pre-trained CNN models for asphalt pavement crack detection and classification. *Automation in Construction*. 160: p. 105297. DOI: https://doi.org/10.1016/j.autcon.2024.105297.

DOI: https://doi.org/10.17509/ijost.v11i1.95376
p- ISSN 2528-1410 e- ISSN 2527-8045