Journal of Theoretical and Applied Information Technology
31st October 2026. Vol.104. No.20
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# LONGITUDINAL CRACK SEVERITY CLASSIFICATION BASED ON CONVNEXT CRACK WIDTH ESTIMATION USING TRANSFER LEARNING

**RATU NURMALIKA[1] , SARIFUDDIN MADENDA[2], NAHDALINA[3], RODIAH[4*]**

[1,3]Lecturer. Gunadarma University, Department of Civil Engineering, West Java, Indonesia

[2]Professor. Gunadarma University, Doctoral Program in Information Technology, West Java, Indonesia

[4]Lecturer. Gunadarma University, Department of Informatics, West Java, Indonesia

E-mail: [1]nurmalikaratu@staff.gunadarma.ac.id, [2]sarif@staff.gunadarma.ac.id, [3]nahdalina@staff.gunadarma.ac.id, [4]rodiah@staff.gunadarma.ac.id,

**ABSTRACT**

Artificial Intelligence (AI) has become an important technology for improving road infrastructure management, particularly in the assessment of pavement damage. Longitudinal cracks are among the most common forms of road deterioration and, if not identified and treated promptly, can reduce transportation safety and increase maintenance costs. Conventional inspection methods rely on manual visual observations, which are often time-consuming, labor-intensive, and prone to subjective errors. This study proposes a transfer learning framework for longitudinal crack severity classification based on crack width estimation using a modified ConvNeXt architecture. The proposed framework consists of dataset preparation, image preprocessing, data augmentation, dataset partitioning, crack width estimation, severity classification, and performance evaluation stages. The ConvNeXt model is modified by replacing the original classification head with a regression layer to predict crack width as a continuous value. The predicted crack width is subsequently mapped into three severity levels, namely light, medium, and heavy cracks. To improve adaptation to the target dataset while preserving pretrained knowledge, a fine-tuning strategy with a low learning rate and Cosine Learning Rate Decay scheduling is employed. Experimental results demonstrate that the proposed model achieves the best performance under a 70:20:10 train-validation-test ratio, obtaining a Test Loss of 625.4404 and a Test $R^2$ score of 0.5050. Furthermore, the average prediction error between manual crack width measurements and model predictions was below 1.1% across all severity categories, indicating strong agreement between estimated and reference measurements. The integration of transfer learning, adaptive learning rate scheduling, and crack width-based severity classification provides an effective and reliable approach for intelligent road condition assessment and decision support in road maintenance management.

**Keywords:** *ConvXNet, Fine Tuning, Learning Rate, Longitudinal Crack, Pipeline*

## 1. INTRODUCTION

Road damage, particularly longitudinal cracking, is one of the most common forms of pavement deterioration in both flexible and rigid pavements [1], [2]. Longitudinal cracks develop parallel to the road direction and may result from thermal expansion, contraction, traffic loading, or structural deficiencies [2], [3]. If not detected and treated at an early stage, such damage can propagate rapidly, reducing road safety, decreasing driving comfort, and increasing maintenance costs [4], [5]. Accurate identification of crack severity is therefore essential for effective pavement management and infrastructure sustainability.

Conventional road crack inspection is primarily performed through manual visual surveys following standardized guidelines such as the FHWA Distress Identification Manual and the Mn/DOT Pavement Distress Identification Manual [6], [7]. Although these procedures provide systematic assessment criteria, they remain labor-intensive, time-consuming, susceptible to subjective interpretation, and difficult to scale for large road networks [8], [9]. Recent advances in artificial intelligence, computer vision, and deep learning have enabled more efficient and objective approaches for pavement distress detection and classification [10]–[12].

Several studies have explored deep learning techniques for crack detection. YOLO-based models have demonstrated real-time detection capabilities for road inspection applications [13], while GAN-based approaches have been utilized to improve image quality and detection performance [14]. Other studies have employed graph-based and attention-based architectures to capture complex crack patterns and improve segmentation accuracy [15], [16]. More recently, ConvNeXt has emerged as an efficient CNN architecture capable of achieving strong feature extraction performance while maintaining computational efficiency [17], [18]. ConvNeXt-based frameworks have been successfully applied to road defect detection, structural crack identification, and industrial surface inspection tasks [19]–[22].

Despite these advances, previous studies still face challenges related to model complexity, computational cost, limited generalization capability, and suboptimal classification performance when encountering crack characteristics that differ from training data [20]–[23]. Furthermore, most studies focus on crack detection or segmentation, while limited attention has been given to the classification of longitudinal crack severity levels. To address this gap, this study proposes a modified ConvNeXt-based classification framework incorporating a pipeline-driven transfer learning strategy. The proposed approach aims to improve feature adaptation, stabilize the fine-tuning process, and enhance classification performance for longitudinal road cracks categorized into light, moderate, and severe classes. The framework includes dataset preparation, preprocessing, data partitioning, model development, crack-width measurement, and performance evaluation stages to support more accurate and reliable road condition assessment.

## 2. METHODS

### 2.1. Road Damage Dataset Acquisition

This study utilized road damage images acquired directly using a vehicle-mounted dashboard camera. Data collection was performed using a Blackvue DR900X-2CH Plus 4K Cloud Dashcam with a resolution of 3840 × 2160 pixels. The camera was installed on the vehicle windshield using a stable mounting system following the recommended installation procedure [24]. The dashcam provides a 162° field of view, enabling broad coverage of the road surface during acquisition. To ensure consistent image quality, the camera was positioned parallel to the road surface and directed toward the center of the driving lane. Data acquisition was conducted along the Jalan Raya Tapos–Depok and Jalan Raya Cikaret–Bogor corridor with a total road length of 14.7 km. Image recording was performed during daytime (12:00–16:00) at a relatively constant vehicle speed of 20–40 km/h following recommended practices for road condition surveys [5]. The acquisition process generated approximately 62 minutes of video footage, as illustrated in Figure 1.




*Figure 1 : Road damage image acquisition using dashcam*

Following data acquisition, video recordings were converted into image frames using Ezgif Video Extractor [25]. Frames were extracted at intervals corresponding to approximately one image every 5 m of vehicle travel to ensure continuous representation of road conditions [26]. Each image was subsequently cropped manually to focus on pavement areas containing longitudinal cracks. Expert annotation was then performed by pavement engineering specialists to determine the severity level of each crack. Longitudinal cracking refers to cracks that develop parallel to the traffic direction and are commonly associated with repeated traffic loading and deterioration of pavement structural layers [27]. Crack severity labeling was conducted according to the ASTM D6433 standard [28], which classifies pavement distress based on crack width measurements. On average, one minute of video recording produced approximately 225 image samples. Examples of the longitudinal crack images used in this study are presented in Figure 2.

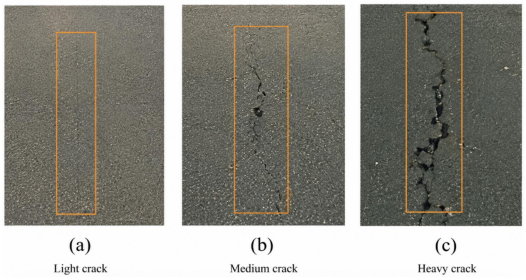



*Figure 2: Examples of longitudinal crack images*

Longitudinal cracks are characterized by several attributes, including crack direction, length, width, depth, continuity, and distribution pattern [2]. Among these characteristics, crack width was selected as the primary feature for severity classification because it is directly related to pavement deterioration and maintenance priority. Previous studies have reported that increasing crack width is associated with greater structural degradation and potential pavement failure [29], [30]. Therefore, the severity levels adopted in this study follow the ASTM D6433 classification standard [28], as summarized in Table 1.

*Table 1 : Severity level based on crack width according to ASTM D6433 [28]*

| Severity Level | Crack Width (mm) |
|---|---|
| Light | ≤13 |
| Medium | >13 and ≤50 |
| Heavy | >50 |

### 2.2. Road Damage Dataset Acquisition

Following dataset acquisition, a preprocessing stage was performed to improve image quality and prepare the data for model training. The preprocessing workflow consisted of four main steps: data cleaning, image augmentation, image resizing, and image normalization. The overall preprocessing procedure is illustrated in Figure 3.

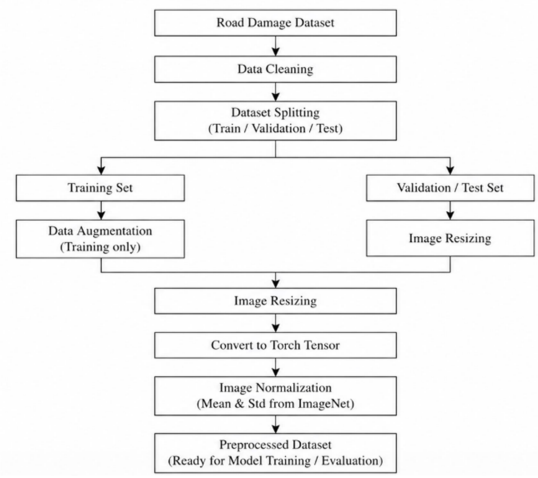



*Figure 3 : Dataset preprocessing stage*

As shown in Figure 3, preprocessing was conducted separately for the training dataset and the validation/testing datasets. The training dataset underwent data augmentation to increase sample diversity and improve model generalization capability. Six preprocessing operations were applied, including horizontal flip, vertical flip, color jitter (brightness and contrast adjustment), rotation, shearing, and image resizing. Data augmentation has been widely adopted to artificially generate new image variations from existing samples, thereby mitigating dataset limitations and reducing the risk of overfitting during model training [31]–[33] . After augmentation, all images were resized to a uniform resolution to ensure compatibility with the ConvNeXt architecture. The processed images were then converted into PyTorch tensor format to enable efficient GPU-based computation [34]. Finally, image normalization was performed using the mean and standard deviation values of the ImageNet dataset. This normalization strategy aligns the input data distribution with the pretrained ConvNeXt weights and contributes to more stable convergence during transfer learning [35].

Unlike the training dataset, the validation and testing datasets were not subjected to augmentation in order to preserve their original characteristics. Only resizing, tensor conversion, and normalization were applied to these datasets. This approach ensures that model evaluation is conducted using unseen and unaltered images, providing a more reliable assessment of classification performance and generalization capability [10].

### 2.3. Image Augmentation

Image augmentation was applied to increase dataset diversity and improve the generalization capability of the deep learning model. This process generates additional image variations from existing samples, reducing the risk of overfitting and enhancing model robustness during training [33], [34]. In this study, five augmentation techniques were employed: horizontal flip, vertical flip, color jitter, rotation, and shearing, as illustrated in Figure 4.

Horizontal and vertical flipping were applied with a probability of 0.5 to introduce orientation variations in crack patterns. Color jitter augmentation was performed using brightness and contrast values of 0.2, allowing the model to better adapt to different illumination conditions encountered during road image acquisition. Rotation augmentation was applied within a range of ±15° to improve the model's ability to recognize cracks captured from different viewing angles. In addition, a shearing transformation of 10° was utilized to simulate geometric distortions that may occur during image acquisition. All augmentation

operations were implemented using the torchvision.transforms library in Python

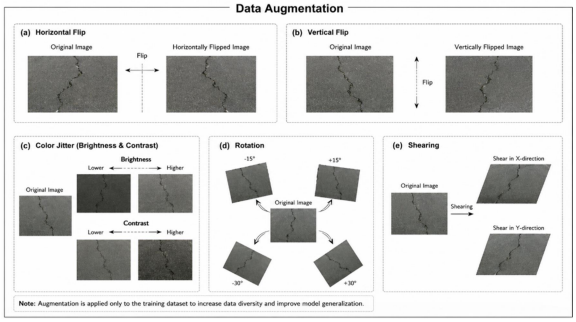



*Figure 4: Augmentation of crack image. (a).horizontal flip; (b).vertical flip; (c) collor jitter; (d).rotation; (e).shearing*

.

Following augmentation, all images were resized to 224 × 224 pixels to ensure compatibility with the ConvNeXt architecture and ImageNet-pretrained weights [17]. The augmentation process substantially increased the number of training samples across all severity classes. Table 2 summarizes the dataset distribution before and after augmentation.

*Table 2: Number of road crack datasets based on three severity classes*

| Number of Datasets | Crack Target Class | | |
|---|---|---|---|
| | Low | Medium | High |
| Groundtruth data | 153 | 194 | 93 |
| After *Cleaning Data* | 107 | 173 | 62 |
| After Augmentation | 535 | 865 | 310 |

The augmented dataset was subsequently used for model training to improve feature learning and classification performance across different levels of longitudinal crack severity.

## 2.4. Image Normalization

Image normalization was performed to standardize the input data distribution and improve model stability during training. First, all images were converted into PyTorch tensor format using the *ToTensor()* function, which transforms image data into a numerical representation suitable for deep learning computation [35]. Subsequently, image normalization was applied using the *Normalize()* function with the ImageNet mean values of [0.485, 0.456, 0.406] and standard deviation values of [0.229, 0.224, 0.225]. These parameters were selected to align the input data distribution with the pretrained ConvNeXt model and facilitate faster convergence during transfer learning [10].

Following normalization, the dataset was organized into mini-batches according to the predefined batch size parameter. Batch-based processing improves training efficiency and enables the model to optimize its parameters iteratively using subsets of the dataset rather than processing the entire dataset at once [10]. To increase sample diversity during training, the data were randomly shuffled before being loaded into each batch. In addition, memory pinning was enabled during the data loading process to improve memory management and accelerate data transfer between the host memory and GPU resources [18]. The resulting normalized dataset was then used as input for the ConvNeXt-based longitudinal crack classification model.

## 2.5. Splitting of Crack Dataset

The augmented dataset, consisting of 535 light, 865 medium, and 310 heavy crack images, was divided into training, validation, and testing subsets using three splitting scenarios: 80:10:10, 70:20:10, and 60:20:20. All experiments were performed using a batch size of 16 and 500 training epochs. Model performance was evaluated using Train Loss, Validation Loss, Train $R^2$, and Validation $R^2$ metrics. As shown in Figure 5, the 80:10:10 split achieved the lowest Validation Loss (113.6032) and the highest Validation $R^2$ (0.6910), outperforming the other configurations. Therefore, the 80:10:10 ratio was selected as the optimal dataset partitioning strategy for subsequent model training and evaluation.

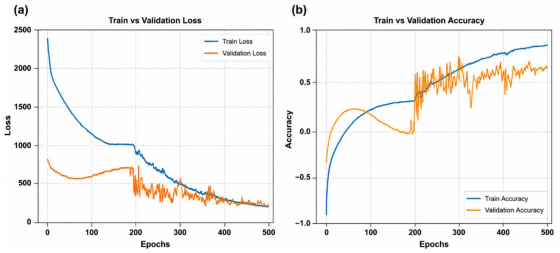



*Figure 5 : Performance Comparison of Different Dataset Splitting Ratios*

## 2.6. Longitudinal Road Crack Classification Model

This study proposes a modified ConvNeXt-based architecture for longitudinal road crack classification. ConvNeXt was selected as the backbone network due to its strong feature

extraction capability, computational efficiency, and effectiveness in image classification tasks [17], [36]. Unlike the standard ConvNeXt architecture, the proposed model incorporates channel attention modules within the residual learning framework to enhance feature representation and improve crack severity discrimination. The model is designed to classify longitudinal road cracks into three severity levels: light, medium, and heavy. The overall architecture of the proposed model is presented in Figure 6.

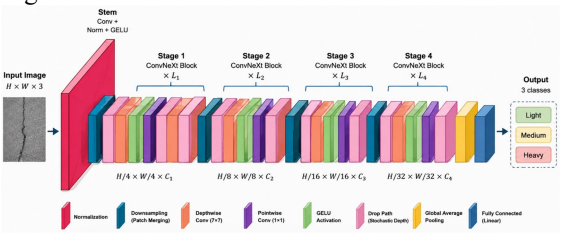



*Figure 6:Modified ConvNeXt model for longitudinal road crack classification*

The model begins with a stem layer that extracts low-level visual features from the input image using convolution operations and the Gaussian Error Linear Unit (GELU) activation function. The extracted features are subsequently processed through multiple ConvNeXt blocks composed of convolution, normalization, residual connection, and activation layers. These blocks enable the network to learn hierarchical feature representations while maintaining training stability and efficient gradient propagation [10].

To improve the identification of crack severity, channel attention mechanisms are integrated into the residual blocks, allowing the network to emphasize informative crack features while suppressing irrelevant background information. Furthermore, residual learning facilitates effective information propagation across deeper layers and preserves discriminative feature representations, which are essential for crack pattern recognition [1], [34]. In the final stage, the extracted feature maps are aggregated using a Global Average Pooling (GAP) layer and forwarded to a fully connected classification layer for severity prediction. The integration of ConvNeXt feature extraction, residual learning, and channel attention enables the proposed architecture to capture crack characteristics more effectively and improve the classification of longitudinal road damage severity.

### 2.7. Crack Width Estimation

To estimate crack width, the ConvNeXt-base architecture was modified by replacing the final classification layer with a single-neuron regression output layer. This modification enables the model to predict continuous crack-width values from road crack images. Prior to inference, all images were resized to 224×224 pixels, converted into tensors, and normalized using the ImageNet mean values [0.485,0.456,0.406][0.485,0.456,0.406][0.485,0.456,0.406] and standard deviation values [0.229,0.224,0.225][0.229,0.224,0.225][0.229,0.224,0.225], consistent with the ConvNeXt pretraining configuration [11]. The normalization process for each color channel was computed using Eq 1 :

$$PixelValue_{Normalized} = \frac{PixelValue - me}{std} \quad (1)$$

where *mean* and *std* represent the ImageNet dataset statistics. This normalization aligns the input distribution with the pretrained model and improves numerical stability during prediction [11]. After preprocessing, the image tensor was fed into the trained model to generate a crack-width prediction in millimeters.

### 2.8. Classification Model Training Pipeline

The proposed longitudinal crack classification model employs a transfer learning strategy to accelerate convergence and improve feature learning performance [37]. Pretrained ConvNeXt weights obtained from the ImageNet dataset were transferred to the feature extraction layers, while the output layer was modified to classify three crack severity levels: light, medium, and heavy.

The training process consisted of two stages. First, an adaptation stage was performed by training only the classification layer using a relatively higher learning rate, enabling the model to adapt to the target dataset while preserving pretrained feature representations [38]. Second, a fine-tuning stage was conducted by unfreezing the network and jointly optimizing both the feature extraction and classification layers using a smaller learning rate to prevent significant modifications to the pretrained weights [39]. To address class imbalance, class weights were computed using Eq. (2):

$$weight = \frac{\sum I}{\sum class * \sum I_c} \quad (2)$$

Where *w* is an equation for calculating weight (*weight*) which is calculated by dividing the number of fracture images ($\Sigma\ I$) by the result of multiplication between the number of longitudinal fracture image classes ($\Sigma\ Class$) and the number of images in *class* ($\Sigma\ Ic$).

Journal of Theoretical and Applied Information Technology
31st October 2026. Vol.104. No.20
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Furthermore, a Cosine Learning Rate Decay scheduler with warmup was employed to adaptively adjust the learning rate throughout training. Model optimization was performed using the selected optimizer and an early stopping strategy, while validation performance was monitored using Loss and ($R^2$) metrics. The overall training workflow of the proposed longitudinal crack classification model is illustrated in Figure 7.

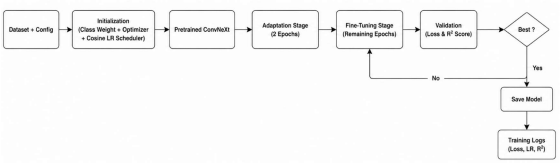



*Figure 7: Pipeline training flow for longitudinal crack classification model*

## 3. RESULT AND DISCUSSION

### 3.1. Longitudinal Crack Classification Results

This study employs a modified ConvNeXt architecture in which the original classification head is replaced with a regression layer that predicts crack width as a continuous value. The predicted crack width is subsequently mapped into three severity categories, namely light, medium, and heavy cracks, based on predefined crack-width thresholds. Table 4 presents the comparison between manually measured crack widths and the crack width predictions generated by the proposed ConvNeXt model, together with the corresponding severity classification results.

*Table 4: Sample Longitudinal Crack Classification Results*

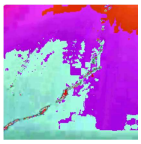
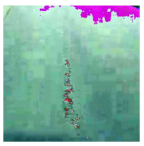
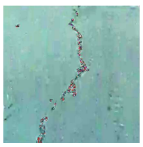
| Normalization Image | Ground Truth | Ground Truth | Ground Truth |
|---|---|---|---|
|  | Light | Light | 98.5 |
|  | Medium | Medium | 96.7 |
|  | Heavy | Heavy | 99.1 |
|  | Heavy | Heavy | 97.8 |
|  | Medium | Medium | 95.4 |

Table 4 presents several examples of longitudinal crack classification results generated by the proposed ConvNeXt-based model. The model successfully classified crack severity into three categories, namely light, medium, and heavy cracks. The predicted labels were highly consistent with the corresponding ground-truth annotations, demonstrating the effectiveness of the proposed transfer learning framework in extracting discriminative crack features. The classification performance was further evaluated under three dataset partitioning scenarios, namely 80:10:10, 70:20:10, and 60:20:20 for training, validation, and testing, respectively. The comparison among these scenarios was conducted to identify the most suitable data distribution for achieving robust and reliable crack severity classification performance.

### 3.2. Testing Performance of the Proposed ConvNeXt Model

Figure 8 illustrates the testing performance of the proposed ConvNeXt-based model throughout the training process. The Test Loss decreases substantially during the initial epochs, indicating that the model progressively learns meaningful representations of longitudinal crack characteristics. After approximately 100 epochs, the loss curve becomes relatively stable, although several fluctuations are still observed, suggesting ongoing adjustments of model parameters during optimization.

The Test $R^2$ curve exhibits a consistent upward trend from negative values in the early training stage toward positive values as training progresses. This behavior indicates an improvement in the model's ability to explain the variability of unseen testing data. The gradual stabilization of the Test $R^2$ curve after the middle training stage further demonstrates that the model achieves a more reliable representation of crack-width features and

severity patterns. Based on the final evaluation, the proposed model achieved a Test Loss of 625.4404 and a Test $R^2$ value of 0.5050. These results indicate that the model was able to reduce prediction errors while explaining approximately 50.5% of the variability in the testing dataset. Although several fluctuations remain during training, the overall trend of decreasing loss and increasing $R^2$ demonstrates the effectiveness of the proposed transfer learning framework in learning discriminative features for longitudinal road crack assessment.

Furthermore, the relatively stable convergence pattern observed in both evaluation metrics indicates that the proposed ConvNeXt architecture successfully learned representative crack features without experiencing severe performance degradation during training. The combination of transfer learning, pipeline-based training, and adaptive learning-rate scheduling contributed to improved model generalization on unseen road crack images. To further validate the robustness of the proposed framework, additional experiments using alternative dataset partitioning strategies were conducted and are discussed in Section 3.4.

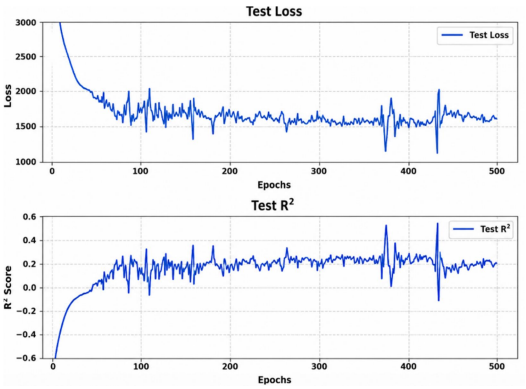



*Figure 8 : Loss and $R^2$ evaluation results of the proposed ConvNeXt model on the testing dataset*

### 3.3. Crack Width Prediction Analysis

Figure 9 presents an example of crack-width prediction generated by the proposed ConvNeXt model for a medium-severity longitudinal crack. The manually measured crack width was 35 mm, whereas the model predicted a width of 36.0893 mm, resulting in an absolute error of 1.0893 mm. This relatively small difference indicates that the proposed model was able to estimate crack width with a high degree of consistency compared with manual measurements. The discrepancy may be influenced by image resolution, feature representation, and the model's generalization capability when processing crack boundaries.

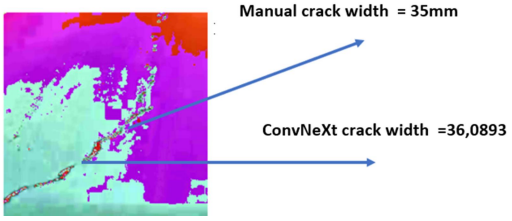



*Figure 9: Example of Image With Difference in Crack Width Measurement Results*

The overall variability between manual measurements and model predictions was further evaluated using the percentage standard deviation shown in Table 5. Lower percentage values indicate that the predicted crack widths were relatively stable and consistent across samples. As shown in Table 5, the heavy crack class achieved the lowest variability (2.903%), followed by the medium class (3.709%) and the light class (4.855%). These results suggest that the proposed model produced more consistent width estimations for larger crack patterns, while prediction variability increased slightly for finer crack structures.

*Table 5 : Standard deviation percentage between manual measurements and ConvNeXt predictions*

| Class | Standard Deviation (%) |
|---|---|
| Light | 4.855 |
| Medium | 3.709 |
| Heavy | 2.903 |

Overall, the low standard deviation percentages demonstrate that the proposed ConvNeXt model provides stable crack-width predictions across different severity levels, supporting its applicability for automated longitudinal crack assessment.

### 3.4. Comparative Analysis of Dataset Partitioning Strategies

To evaluate the effect of dataset partitioning on model performance, additional experiments were conducted using three train-validation-test ratios, namely 80:10:10, 70:20:10, and 60:20:20. The objective was to determine the most suitable data distribution for longitudinal crack classification. Figure 10 presents the testing performance obtained from the 70:20:10 and 60:20:20 partitioning schemes, while the results of the selected 80:10:10 configuration were previously presented in Figure 8. As shown in Figure 10(a), the 70:20:10 ratio

produced a Test Loss of 1283.8495 and a Test $R^2$ value of 0.2675. The relatively high loss and low coefficient of determination indicate that the model was only able to explain approximately 26.75% of the variability in the testing data. This result suggests limited generalization capability, where the learned feature representations were not sufficiently robust when applied to unseen crack images. Figure 10(b) shows the performance of the 60:20:20 partitioning scheme, which achieved a Test Loss of 639.2193 and a Test $R^2$ value of 0.3875. Compared with the 70:20:10 configuration, this scenario reduced the prediction error and improved the model's explanatory power. However, the achieved performance remained lower than that obtained using the 80:10:10 ratio.

Among all evaluated scenarios, the 80:10:10 partitioning strategy achieved the best overall performance, with the lowest Test Loss (625.4404) and the highest Test $R^2$ value (0.5050). These findings indicate that allocating a larger proportion of data for training enabled the model to learn more representative crack features, resulting in improved prediction accuracy and generalization performance. Therefore, the 80:10:10 ratio was selected as the optimal dataset partitioning strategy for the proposed ConvNeXt-based longitudinal crack classification framework.

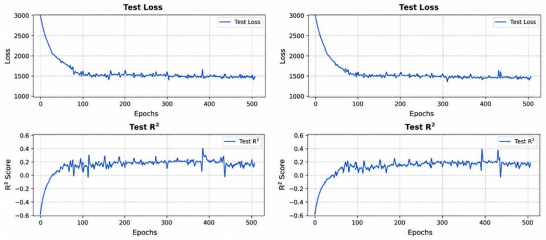



*Figure 10 : Testing Performance Comparison Under Different Dataset Partitioning Strategies: (a) 70:20:10 Ratio, and (b) 60:20:20 Ratio.*

### 3.5. Performance Comparison with Recent ConvNeXt-Based Road Crack Studies

Table 6 compares the performance of the proposed ConvNeXt modification with several recent studies that employed ConvNeXt-based architectures for crack detection and classification tasks. The comparison demonstrates that the proposed approach achieved an overall classification accuracy of 91% for three longitudinal crack severity classes (light, medium, and heavy), outperforming several previous ConvNeXt-based methods.

*Table 6 : Comparative Analysis of Dataset Partitioning Strategies*

| Study | Method | Performance |
|---|---|---|
| Shi et al. (2024) [18] | CBU-FI Net (U-Net + ConvNeXt) | Accuracy: 82.47%; CrackSeg9k: 81.38% |
| Li et al. (2023) [19] | CL-YOLOv8 with ConvNeXtV2 Backbone | Accuracy: 85.3%; mAP: 83.7% |
| Wu et al. (2023) [20] | DadNet with ConvNeXt Feature Extractor | MIoU: 78.95%; Jaccard: 61.05% |
| Wang et al. (2023) [21] | Modified ConvNeXt-Based YOLOv7 | mAP: 82.9% |
| Proposed Model | Modified ConvNeXt with Transfer Learning Pipeline | Accuracy: 91.0% |

Shi et al. reported classification accuracies of 82.47% and 81.38% using the CBU-FI Net architecture that combines U-Net and ConvNeXt for microcrack detection. Similarly, Li et al. achieved an accuracy of 85.3% using the CL-YOLOv8 framework that integrates ConvNeXt V2 as the backbone network. Wang et al. reported a mean Average Precision (mAP) of 82.9% through modifications of the ConvNeXt backbone and pooling layers, while Wu et al. obtained an MIoU score of 78.95% using a ConvNeXt-based attention network.

The superior performance achieved by the proposed model can be attributed to the integration of pipeline-based transfer learning, adaptive learning-rate scheduling, and crack-width-based severity classification. Unlike previous studies that mainly focused on crack detection or segmentation, the proposed framework directly addresses longitudinal crack severity classification by combining crack-width estimation and feature learning within a unified ConvNeXt architecture. This strategy enables the model to learn more discriminative crack representations while maintaining stable convergence during training. Overall, the results indicate that the proposed ConvNeXt modification with pipeline flow provides a competitive and effective solution for automated longitudinal road crack assessment, achieving higher classification performance than

several recently published ConvNeXt-based approaches.

## 4. CONCLUSION

This study proposed a modified ConvNeXt-based framework for longitudinal road crack severity classification using crack-width estimation and transfer learning. The proposed approach successfully classified longitudinal cracks into three severity levels (light, medium, and heavy) according to the ASTM D6433 standard by utilizing predicted crack width as the primary classification feature. Experimental results demonstrated that the integration of transfer learning, residual learning, and channel attention mechanisms enabled the model to effectively learn discriminative crack characteristics from road pavement images. Among the evaluated dataset partitioning strategies, the 80:10:10 ratio achieved the best performance, yielding a Test Loss of 625.4404 and a Test $R^2$ value of 0.5050, indicating the lowest prediction error and the highest capability to explain variability in unseen test data. These findings demonstrate that the proposed ConvNeXt modification provides a robust and effective solution for automated longitudinal road crack severity assessment and has strong potential for supporting intelligent road monitoring and pavement maintenance systems.

## REFERENCES:

[1] D. Gu, S. Ou, and G. Liu, "Global burden of road injuries and their attributable risk factors from 1990 to 2021: A systematic analysis for the Global Burden of Disease Study 2021," *Preventive Medicine Reports*, vol. 53, p. 103051,2025.DOI:https://doi.org/10.1016/j.pmedr.2025.103051.

[2] K. Chen, L. Ji, Z. Y. Xiao, J. Q. Wu, L. L. Xu, and Z. C. Cheng, "Correlation analysis of longitudinal cracks and vertical deformation within asphalt pavement of cold regions,"*Research in Cold and Arid Regions*, vol. 15, no. 6,pp.278–287,2023,DOI: 10.1016/j.rcar.2023.11.001.

[3] D. X. Xiao and Z. Wu, "Longitudinal cracking of jointed plain concrete pavements in Louisiana: Field investigation and numerical simulation," *International Journal of Pavement Research and Technology*, vol. 11, no. 5, pp. 417–426, 2018, DOI: 10.1016/j.ijprt.2018.07.004.

[4] A. M. Mosa, L. A. Salem, and Q. S. Banyhussan, "Treatment of cracking in rigid highway pavements using knowledge-based system," *International Journal of Innovation in Engineering*, vol. 2, no. 1, pp. 68–77, Feb. 2022.
DOI: https://doi.org/10.59615/ijie.2.1.68

[5] World Health Organization, *Global Status Report on Road Safety 2023*. Geneva: WHO, 2023.

[6] M. Alzarrad et al., "Deep learning-based pavement distress detection and classification: A review of recent advances," *Automation in Construction*,vol.147,2023.DOI:https://doi.org/10.1016/j.autcon.2023.104637.

[7] S. Gopalakrishnan et al., "AI-based pavement condition assessment and maintenance decision systems: A systematic review," *Journal of Infrastructure Systems*, vol. 29, no. 4, 2023. DOI: https://doi.org/10.1061/(ASCE)IS.1943-555X.0000721

[8] P. M. S. Raihan et al., "Pavement Crack Detection and Solution with Artificial Intelligence," *European Journal of Theoretical and Applied Sciences*, vol. 2, no. 4, pp. 277–314, 2024, DOI: 10.59324/ejtas.2024.2(4).25.

[9] I. Meftah et al., "Visual Detection of Road Cracks for Autonomous Vehicles Based on Deep Learning," *Sensors*, vol. 24, no. 5, 2024, DOI: 10.3390/s24051647.

[10] L. Alzubaidi et al., "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data*, vol. 8, no. 1, 2021, DOI: 10.1186/s40537-021-00444-8.

[11] B. Kulambayev et al., "A Deep Learning-Based Approach for Road Surface Damage Detection,"*Computers, Materials and Continua*, vol.73,no.2,pp.3403-3418, 2022, DOI:10.32604/cmc.2022.029544.

[12] W. S. Qureshi et al., "Deep learning framework for intelligent pavement condition rating: A direct classification approach for regional and local roads," *Automation in Construction*, vol. 153, p. 104945, 2023, DOI: 10.1016/j.autcon.2023.104945.

[13] R. Wang et al., "Development of an Improved YOLOv7-Based Model for Detecting Defects on Strip Steel Surfaces," *Coatings*, vol. 13, no. 3, 2023, DOI : 10.3390/coatings13030536.

[14] T. Zhang, L. Qin, Q. Zou, L. Zhang, R. Wang, and H. Zhang, "CrackScopeNet: A Lightweight Neural Network for Rapid Crack Detection on Resource-Constrained Drone Platforms,"

*Drones*, vol. 8, no. 9, p. 417, 2024, DOI: 10.3390/drones8090417.

[15] M. Wu, R. Chai, Y. Zhang, and Z. Lu, "Dadnet: Dual-attention detection network for crack segmentation on tomb murals," *Heritage Science*, vol. 12, no. 1, pp. 1–13, 2024, DOI: 10.1186/s40494-024-01474-0.

[16] R. Duan, Z. Yan, P. Huang, S. Li, and Q. Hu, "A Method for Steel Crack Detection Based on Multispectral Image Fusion," *SSRN Electronic Journal*, 2021, DOI: 10.2139/ssrn.4556865.

[17] M. Tan and Q. V. Le, "EfficientNetV2: Smaller Models and Faster Training," *Proceedings of the 38th International Conference on Machine Learning (ICML)*, 2021. DOI:https://doi.org/10.48550/arXiv.2104.00298

[18] Y. Zhang et al., "A survey on efficient deep learning for computer vision: model compression, acceleration and neural architecture search," *IEEE Access*, vol. 10, pp. 1–25,2022. DOI: https://doi.org/10.1109/ACCESS.2022.3152810

[19] H. Zhang, C. Liu, J. Ho, and Z. Zhang, "Crack Detection Based on ConvNeXt and Normalization," *Journal of Physics: Conference Series*, vol. 2289, no. 1, 2022, DOI: 10.1088/1742-6596/2289/1/012022.

[20] Z. Zhou, W. Zhao, J. Li, and K. Shi, "SPCNet: A Strip Pyramid ConvNeXt Network for Detection of Road-Surface Defects," *Signal, Image and Video Processing*, vol. 18, pp. 37–45, 2024, DOI: 10.1007/s11760-023-02698-6.

[21] Q. Li, G. Zhang, and P. Yang, "CL-YOLOv8: Crack Detection Algorithm for Fair-Faced Walls Based on Deep Learning," *Applied Sciences*, vol. 14, no. 20, 2024, DOI: 10.3390/app14209421.

[22] C. Shi et al., "Deep learning-based crack segmentation for structural health monitoring of bridges using convolutional neural networks," *Automation in Construction*, vol. 145, 2023. DOI:https://doi.org/10.1016/j.autcon.2022.104598

[23] C. Shi et al., "Real-Time ConvNeXt-Based U-Net with Feature Infusion for Egg Microcrack Detection," *Agriculture*, vol. 14, no. 9, 2024, DOI: 10.3390/agriculture14091655.

[24] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015. DOI: https://doi.org/10.1038/nature14539.

[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017. DOI: https://doi.org/10.1145/3065386.

[26] Z. Chen, Y. Luo, J. Wang, J. Li, C. Wang, and D. Li, "DPENet: Dual-path extraction network based on CNN and transformer for accurate building and road extraction," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 124, p. 103510, 2023.

[27] X. Yang et al., "Automation in road distress detection, diagnosis and treatment," *Journal of Road Engineering*, vol. 4, no. 1, pp. 1–26, 2024.

[28] ASTM International, "ASTM D6433-20: Standard Practice for Roads and Parking Lots Pavement Condition Index Surveys," ASTM International, 2020.

[29] Wisnumurti, A. Soehardjono, and R. M. Simatupang, "Effect of variations in concrete quality on the crack width in rigid pavement," *Eastern-European Journal of Enterprise Technologies*, vol. 1, no. 1(127), pp. 33–40, 2024. DOI: https://doi.org/10.15587/1729-4061.2024.298680.

[30] Z. Liu, S. Yu, Y. Huang, L. Liu, and Y. Pan, "A systematic review of rigid-flexible composite pavement," *Journal of Road Engineering*, vol. 4, no. 2, pp. 203–223, 2024. DOI:https://doi.org/10.1016/j.jreng.2024.02.001

[31] P. P. Martin and N. Graulich, "Navigating the Data Frontier in Science Assessment: Advancing Data Augmentation Strategies for Machine Learning Applications with Generative Artificial Intelligence," *Computers and Education: Artificial Intelligence*, vol. 7, p. 100265, 2024.DOI:https://doi.org/10.1016/j.caeai.2024.100265

[32] A. Tatar, M. Haghighi, and A. Zeinijahromi, "Experiments on image data augmentation techniques for geological rock type classification with convolutional neural networks," *Journal of Rock Mechanics and Geotechnical Engineering*, vol. 17, no. 1, pp. 106–125, 2025. DOI:https://doi.org/10.1016/j.jrmge.2024.02.015

[33] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, article 60, 2019. DOI: https://doi.org/10.1186/s40537-019-0197-0

[34] P.Padmapoorani and S.Senthilkumar, "Application of machine learning for crack detection on concrete structures using CNN architecture," *Matéria (Rio de Janeiro)*, vol. 28, no. 1, 2023.DOI: https://doi.org/10.1590/1517-7076-rmat-2023-0010

[35] T. Wang et al., "Understanding and improving normalization techniques in deep convolutional neural networks for vision tasks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 4, pp. 1–12, 2024. DOI:https://doi.org/10.1109/TNNLS.2024.3351234

[36] L. Q. Trung, N. Kasai, M. Le, and K. Sekino, "Predicting actual crack size through crack signal obtained by advanced flexible eddy current sensor using ResNet integrated with CBAM and Huber loss function," *NDT & E International*, vol. 149, p. 103249, 2025. DOI:https://doi.org/10.1016/j.ndteint.2024.103249

[37] A. Kirillov et al., "Segment Anything," *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023. DOI:https://doi.org/10.1109/ICCV51070.2023.00045

[38] R. Surendran, J. Anitha, and D. J. Hemanth, "Inside Feature Learning for Image Classification Using Transfer Learning Approach," *Frontiers in Artificial Intelligence and Applications*, vol. 347, pp. 447–455, 2022. DOI: https://doi.org/10.3233/FAIA220051

[39] 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.DOI: https://doi.org/10.3390/s21051688