Utilising Convolutional Neural Networks for Pavement Distress Classification and Detection.

Mhd Saeed Sharif
Computer Science and DT, ACE, UEL
University Way, London, UK
S.Sharif@uel.ac.uk

Damilola Ibukun Emiola
Computer Science and DT, ACE, UEL
University Way, London, UK
u2234923@uel.ac.uk

Abin Zorto
Computer Science and DT, ACE, UEL
University Way, London, UK
u2091940@uel.ac.uk

Alex Apeagyei
Department of Engineering and Construction, ACE, UEL, London UK
a.apeagyei@uel.ac.uk

Wael Elmedany
College of Information Technology
University of Bahrain, Kingdom of Bahrain
welmedany@uob.edu.bh

***Abstract*—This paper examines deep learning models for accurate and efficient identification and classification of pavement distresses. In it, a variety of related studies conducted on the topic as well as the various identification and classification methods proposed, such as edge detection, machine learning classification informed by statistical feature extraction, artificial neural networks, and real-time object detection systems, are discussed. The study investigates the effect of image processing techniques such as grayscaling, background subtraction, and image resizing on the performance and generalizability of the models. Using convolutional neural networks (CNN) architectures, this paper proposes a model that correctly classifies images into five pavement distress categories, namely fatigue (or alligator), longitudinal, transverse, patches, and craters, with an accuracy rate of 90.4% and a recall rate of 90.1%. The model is contrasted to a current state-of-the-art model based on the You Only Look Once framework as well as a baseline CNN model to demonstrate the impact of the image processing and architecture building techniques discussed on performance. The findings of this paper contribute to the fields of computer vision and infrastructure monitoring by demonstrating the efficacy of convolutional neural networks (CNNs) in image classification and the viability of using CNN-based models to automate pavement condition monitoring.**

***Index Terms*—Pavement Condition Monitoring, Computer Vision, Image Processing, Convolutional Neural Networks, Object Detection.**

I. Introduction

Over the past few decades, the field of computer vision has evolved from studies on the brain's visual neurons as conducted by Harvard neurophysiologists Hubel and Wiesel [1] to its practical applications in various industries and fields of study today. The work of Hubel and Wiesel in 1959 explored how the visual system builds an image from simple stimuli into more complex representations by observing a kitten's visual cortex neural responses to series of images. This research indicated that image processing begins with the recognition of small shapes and patterns found in images.

In the mid-1970s, the introduction of Optical Character Recognition (OCR) and Intelligent Character Recognition (ICR) technology further propelled the study of Artificial Intelligence [2]. Fukushima leveraged the findings from preceding studies and proposed a neural network model called the "neocognitron" in 1980. The neocognitron was a self-organized neural network that could learn without a teacher and recognize stimulus patterns [2].

These discoveries eventually led to the development of convolutional neural networks (CNNs) which include a convolutional layer for feature extraction and subsequent layers for image classification. Lecun and Bengio [3] proposed the use of multi-layer back-propagation to turn the initial layers in neural networks into feature extractors.

This research focuses on investigating the effectiveness of convolutional neural network (CNN)-based image processing and object detection tools in accurately identifying and categorizing pavement cracks. The study aims to develop a reliable tool for automated condition assessment of road surfaces and explore the feasibility of implementing real-time classifiers for pavement crack detection.

The research also addresses subject-specific factors that can impact classifier performance and examines image processing techniques to improve prediction accuracy. It analyzes different types of pavement distress, their causes, and the consequences of not identifying and addressing them promptly. The study aims to identify potential barriers to the implementation of CNN-based pavement condition monitoring systems in real-world scenarios and propose solutions to overcome them. It seeks to answer research questions related to real-time detection and classification of pavement distress, image processing techniques, and challenges encountered during data collection and model building.

The research objectives include developing and training a CNN for accurate detection and classification of pavement distress, conducting a comparative analysis with pre-defined data detection tools, investigating challenges and limitations, exploring transfer learning techniques, and assessing scalability and computational efficiency.

Expected outcomes include the creation of a model for accurate pavement crack classification, improving efficiency and accuracy in pavement condition monitoring, early detection and timely intervention to prevent further deterioration and safety hazards, optimization of existing infrastructure, and cost savings.

Overall, the research aims to contribute to the field of automated pavement monitoring and provide a foundation for advancements in pavement crack detection technologies.

II. Literature Review

A. Image Processing

Image processing plays a crucial role in the quality of results obtained from image classification models. [4] emphasize the significance of image processing in improving the accuracy of crack detection. They propose adaptive thresholding to enhance the contrast between distresses and the background, as cracks tend to have darker contours. Applying thresholding techniques can improve a model's ability to identify crack features in input data. The authors also highlight the need for filtering to remove noise, deblur images, and highlight specific crack features.

Data collection methods can greatly impact the quality of input data for image classification models. Radopoulou and Brilakis [5] propose the use of vision trackers installed in civilian vehicles to automatically detect patches in video frames, providing a cost-effective alternative to manual data collection. They emphasize the importance of image processing techniques such as grayscale transformation, median filtering, histogram equalization, image binarization, and morphological operations in improving the quality of training data. They suggest that pavement distress detection requires attention to detail, as these distresses can be challenging to detect even for the human eye. The authors also highlight the importance of addressing blurriness in images through techniques like median filtering and the use of grayscale transformation to reduce noise associated with color.

Balbin et al. [6] propose image processing techniques such as Hough transform, dilation, grayscale transformation, and Haar wavelet transform as precursors to automatic surface crack detection. They use edge detection to isolate crack contours for better detection. Surface crack classification is performed using a Haar trained cascade object detector trained on positive and negative samples.

Liu et al. [7] employ image grayscaling and smoothing techniques to reduce noise introduced by color and blurriness. They apply binary image denoising based on the 8-neighborhood feature of noise to eliminate isolated noise points. The authors study the trajectory, length, width, and area of specific crack types as features for classification. Length and width are extracted through the distance between image pixels, while the crack area is calculated as the product of length and width.

B. Image Classification

Different authors have investigated various approaches for the object detection/classification portion of these tasks, including edge detection, feature extraction for machine learning, and deep learning.

1) Edge Detection: Nienaber et al. [8] conducted a study on the use of edge detection to identify potholes in South African roads. They utilized the Canny edge detection algorithm to identify and highlight the distinctive edges of potholes in the captured images. The assumption was that the darker lining along the edges of a pothole, in contrast to the rest of the road, would serve as a key feature for detection.

However, this approach may have limitations. It fails to consider the possibility that some potholes may be obscured by dust or debris, causing their edges to blend with the surrounding pavement and making them indistinguishable through edge detection alone. Additionally, the research primarily focused on detecting a single distress class (potholes), which may limit its effectiveness in detecting other types of pavement distress, such as longitudinal, transverse, or alligator cracks. Different distress types may have distinct characteristics that require tailored detection approaches.

Nevertheless, the edge detection approach has its applications in other areas of research. Dwivedi et al. [9] utilized edge detection techniques, including smoothing, intensity gradient calculation, non-maximum suppression, and hysteresis thresholding, to detect edges in images for the purpose of identifying unattended safety breaches in public spaces such as parks, shopping malls, and transportation hubs. This research aimed to enhance security measures by promptly and accurately identifying potential security threats.

2) Machine Learning: Shi et al. [10] focused on the traditional approach of feature extraction from crack images and employed machine learning techniques for detection. They initially used mean and standard deviation values as features and then incorporated channel features such as a histogram of oriented gradient to achieve faster detection results. By combining different colors, magnitudes, and orientation channels, they extracted a total of 3328 distinct features.

To address the issue of overfitting in decision trees, they utilized a random structured forest, which is an ensemble of randomized decision trees. This ensemble approach helps mitigate the overfitting problem by training multiple models that contribute to the final result.

One noteworthy aspect of their work is their cautious approach to image binarization. They avoided traditional thresholding methods that tend to classify small regions as noise based on their size, potentially removing inconspicuous cracks. Instead, they proposed a crack descriptor that utilizes statistical features of structured tokens to identify unique properties of cracks and distinguish them from noise.

However, a limitation of their research is the inadequate consideration of large volumes of noise typically found on busy highways, where pavement condition monitoring is most crucial. The training images used in their study do not include non-pavement objects in the background, which could be a nuisance if detected in new data.

Other researchers, such as Gavilan et al. [11], Nguyen and Hoang [12], Hoang et al. (2018), and Zhang et al. [13], have also proposed machine learning classifiers, including support vector machines and random forest classifiers, for the detection of pavement cracks. These studies contribute to the exploration of different machine learning approaches for crack detection and classification.

3) Convolutional Neural Networks: Convolutional neural networks (CNNs) have gained popularity in recent years for classification of images with patches. Zhang et al. [14] trained a CNN model for crack patch classification, outperforming support vector machines and tree-based models. Chu et al. [15] proposed a pothole crack detection (PCD) model based on CNNs optimized with K-fold cross-validation, achieving high precision and recall rates in identifying potholes. Pintelas et al. [16] introduced Multiview convolutional neural networks, which enhanced the performance of pre-trained neural networks like ResNet and VGG. Cha et al. [17] utilized CNNs for concrete crack detection, aiming to create a model that is robust to noise introduced by lighting, shadow casting, and blurriness.

Issa et al. [18] focused on predicting pavement condition indexes using artificial neural networks (ANNs), designing a multilayer architecture with input neurons receiving inputs related to pavement conditions. Singh et al. [19] developed an anomaly recognition software based on CNNs, processing CCTV footage and using a pre-trained network for classification. Mandal et al. [20] applied deep learning techniques to predict irregularities in road pavements, employing a You Only Look Once (YOLO) v2-based deep CNN.

Huyan et al. [21] proposed the Crack Deep Network (CrackDN) based on the Faster Region Convolution Neural Network (F-RCNN) architecture, which achieved effective crack detection in complex backgrounds. Song et al. [22] compared the performance of F-RCNNs with traditional CNNs and a K-value method, finding that the F-RCNN ensemble performed better in detecting pavement distress.

Melegrito et al. [23] developed a YOLO v3-based transfer learning CNN model to detect abandoned carts, while Manalo et al. [24] trained a transfer learning model based on the YOLO v3 framework using the ImageAI python library. These studies demonstrate the application of CNNs and related architectures in various image classification tasks, including crack detection, pothole detection, and object detection in parking lots.

III. Methodology

This study aims to compare two different approaches for detecting and classifying pavement distress: the proposed CNN-based approach and the YOLO v8 approach. The flow of methodology components is depicted in Figure 1.

A. Data Collection

To gather the necessary data for this study, we utilized the Roboflow's Pavement Distresses v2 Image Dataset. This dataset contains a wide range of images showcasing road pavements in various states of distress. Prior to conducting this study, the dataset was annotated using TensorFlow to identify the bounding boxes, and the annotations were saved in a CSV document. The dataset includes a total of 680 images, with 1009 bounding boxes and annotations. Some images have multiple annotated bounding boxes.

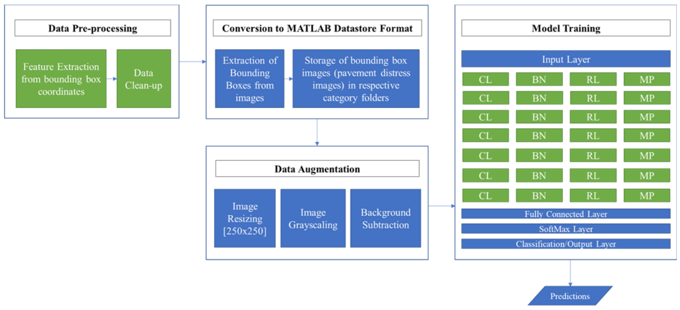

Fig. 1. Flow of Methodology Components

B. Dataset Description

This study utilizes image data sourced from Roboflow's Pavement Distresses v2 Image Dataset. Roboflow is a platform that hosts a vast collection of open-source computer vision datasets and APIs across various fields and industries. The dataset used in this study consists of images depicting road pavements in different states of distress. Prior to this study, object detection was performed on the dataset using TensorFlow, and the annotations were saved in a CSV document. The dataset was then divided into train, test, and validation folders.

The dataset comprises a total of 680 images with 1009 bounding boxes and annotations. Some images contain more than one annotated bounding box. The annotations fall into five distress categories, which are described below.

1) Fatigue (or Alligator) Cracking: Fatigue cracking is a common type of pavement distress characterized by interconnected cracks caused by load-related deterioration. It is prevalent in areas with high traffic volumes, occurring when the asphalt layer of pavements is insufficient to support repeated traffic loads. Longitudinal cracks form along the wheel paths and eventually connect, resembling the back of an alligator or crocodile. Fatigue cracking leads to roughness on the pavement surface and allows moisture infiltration, which can result in potholes and pavement disintegration if not detected and treated early. There are 172 images in this category (see Figure 2).

2) Transverse Cracking: Transverse cracks occur perpendicular to the pavement's centerline. They also lead to moisture infiltration and roughness, similar to block cracks. These cracks are usually caused by thermal cracking, which involves the shrinkage of the asphalt surface due to low temperatures or asphalt binder hardening. However, factors such as inadequate pavement thickness or excessive traffic loads can contribute to their development as well. The dataset contains 246 images in this category (see Figure 3).

3) Longitudinal Cracking: Longitudinal cracks occur parallel to the pavement's centerline and indicate possible fatigue cracking and structural failure. They also lead to moisture infiltration and roughness. These cracks can be caused by poorly constructed or located joints, reflective cracks from underlying layers, asphalt fatigue, or top-down cracking. The

Fig. 2. Collage of Images in the Fatigue Category

Fig. 4. Collage of Images in the Longitudinal Crack Category

Fig. 3. Collage of Images in the Transverse Crack Category

Fig. 5. Collage of Images in the Pothole Category

dataset includes 231 images in this category (see Figure 4).

4) Potholes: Potholes are small bowl-like depressions that penetrate the pavement surface down to the base course. They occur most frequently on roads with thin HMA (Hot Mix Asphalt) surfaces and can cause severe roughness and vehicular damage or accidents, particularly on highways. Potholes also lead to moisture infiltration, further deteriorating the pavement. They generally form as a result of severe fatigue cracking, where interconnected cracks create small chunks of pavement that become dislodged. The dataset contains 155 images of potholes (see Figure 5).

5) Patching: Patching refers to corrective measures taken to repair localized pavement distresses. These patches appear darker than the surrounding pavement area due to their relative youth. Although patches may perform well, they contribute to overall pavement roughness. The only way to remove patches is by applying a structural or non-structural overlay to the entire pavement area. The dataset contains 205 images of patches (see Figure 6).

C. Dataset Summary

The dataset comprises images categorized into Fatigue, Longitudinal, Transverse, Patch, and Pothole groups. The images have varying dimensions and were captured using colored lenses. Some misclassifications are present, such as patches mistakenly categorized as potholes. Additionally, the alignment of transverse and longitudinal crack images may be inaccurate due to portrait mode instead of landscape mode, causing them to appear parallel rather than perpendicular. Some images from both categories may appear diagonal, lacking a clear point of reference for determining their alignment with the pavement's centerline. A few images also include objects other than pavement distress, such as vehicles and traffic cones.

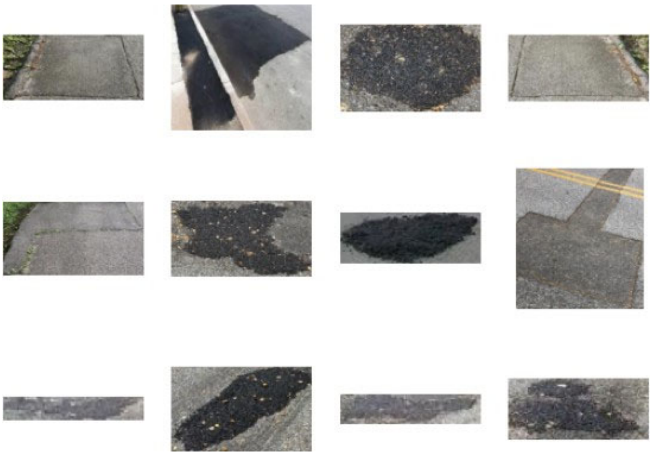

Fig. 6. Collage of Images in the Patches Category

D. Data Analysis Methods

In this section, we discuss the analysis conducted in this research, which involved data pre-processing and model architecture. The focus was on training a CNN-based model using the MATLAB integrated development environment (IDE).

1) Preprocessing: During the pre-processing phase, the data from the Roboflow database was obtained in image and CSV format (Roboflow). The three image folders (train, test, and val) along with their corresponding annotation files were merged and uploaded to Google Drive for further processing using the Python IDE in Google Colab.

To address an issue of smaller cracks being classified as noise and affecting model performance, the data cleaning process started by removing tiny images with dimensions below 20 pixels. This was achieved using the pandas library, which allowed for subsetting the data based on specific criteria. The height and width of the bounding boxes were extracted from the annotation files and used to create two new columns in the DataFrame. Images that did not meet the minimum size requirement were eliminated as they contained minimal information and introduced noise to the model.

Additionally, images predominantly classified as transverse or longitudinal cracks were adjusted to match their correct orientation. By comparing the height and width of the bounding boxes, longitudinal cracks were reclassified as transverse cracks if the height was greater than the width, and vice versa. However, diagonal fractures within these categories were not addressed in this reclassification process.

The os library played a crucial role in navigating the file system (os). It provided functions for creating, renaming, moving, and deleting files and directories, as well as retrieving information about the file system and system-related details. In this study, the os library was used to extract file paths, file names, and folder names for further analysis.

The cv2 library, which specializes in image and video processing, was extensively used in this study (cv2). It offered functions for reading and writing image files, image filtering, thresholding, and image resizing. Specifically, the imread function was used to read each file, and the bounding box coordinates from the annotation files were used to extract the detected cracks. The chdir function from the os library facilitated switching between directories, allowing the extracted images to be saved in the respective category folders using the imwrite function from cv2

E. Model Building

The analysis involved using the MATLAB IDE for the deep learning part. The data was cleaned and uploaded to MATLAB Drive, then read into the IDE using the imageData-store function. The folder names were used as labels for the images. The data was split into train and validation datasets, with an 80:20 ratio. The labels for the validation dataset were saved for later evaluation. To augment the datasets, the augmentedImageDatastore function was used to resize the images to 250 x 250 and remove background images to reduce noise.

Alternatively, the MATLAB transform function could have been used for applying various image processing techniques like binarization, resizing, grayscaling, and histogram equalization. However, the transform function's output couldn't be directly used with the trainNetwork function, so additional steps would be required to combine the transformed data with the image labels.

The network architecture consisted of an input layer, 30 hidden layers, and an output layer. The hidden layers included convolutional, batch normalization, relu, and max pooling layers. The convolutional layers performed feature extraction by applying convolution operations to detect patterns in the input data. The batch normalization layers standardized the input data to improve performance and mitigate overfitting. The relu layers applied the rectified linear unit activation function to learn complex features. The max pooling layers reduced the spatial dimensions of the feature maps. Finally, the fully connected layer connected every neuron in the input to every neuron in the output, and the softmax layer provided probability distributions over the classes.

The classificationLayer was used as the output layer to generate probability distributions and make predictions based on the highest probability. This layer ensured the outputs were normalized and could be interpreted as probabilities.

F. Limitations

The proposed CNN-based method for detecting pavement distresses has several limitations. Firstly, the model's inference time was found to be slow, especially when processing input images. This slow performance would be further exacerbated when working with larger datasets with higher dimensions. Consequently, the proposed method may not be suitable for real-time pavement distress detection scenarios where prompt results are crucial. Additionally, the MATLAB environment used in the study lacked sufficient resources for effective troubleshooting of potential implementation issues. This limitation could introduce challenges in ensuring the smooth functioning and reliability of the proposed method. Given these limitations, it is recommended to explore alternative approaches, such as model training in a Python IDE, or to optimize the CNN-based method to improve inference speed. Additionally, addressing the need for better troubleshooting resources and comprehensive documentation within the MATLAB environment would be beneficial.

IV. RESULTS

In MATLAB, these images were divided into two datastores using an 80:20 division, with 80% of the data set aside for training the model and 20% for validating and evaluating its performance. At the conclusion of this research, the highest validation set accuracy achieved was 89.90%.

TABLE I
RESULTS FROM THE PROPOSED MODEL

CLASS	INSTANCES	PRECISION	RECALL
Fatigue	34	0.813	0.765
Longitudinal	45	0.795	0.867
Patch	42	0.953	0.976
Pothole	28	1.000	1.000
Transverse	49	0.957	0.898
All	198	0.904	0.901

While the model accurately predicted the majority of images in the validation dataset, it had trouble distinguishing between fatigue fracture, longitudinal crack, and transverse crack images. Six fatigue fractures were predicted to be longitudinal, four transverse cracks were predicted to be longitudinal, and four longitudinal cracks were predicted to be transverse. In addition, there were a few minor misclassifications of fatigue and longitudinal fractures as patches and transverse cracks, as well as a misclassification of a single patch and transverse crack as fatigue cracks. This category was not misclassified in any way (potholes were not misclassified as other categories or other categories were not misclassified as potholes). It can be inferred from these results that Nienaber et al's [8] hypotheses are accurate, as the model was able to distinguish craters from other study categories. This may be due to the difference in pothole outlines as they appear in images or the width of the contrast between the distress and the remainder of the pavement surface. This model's parameters were modified and optimised for optimal performance. The options that produced the greatest results were solver: Adam., The Learning Rate is 0.01 and the Validation Frequency is 40.

Using the same image dataset, YOLOv8 baseline and pre-trained models were also examined for comparison. Both models had 225 layers, were trained with 570 images, and were evaluated with 54 images from a validation dataset. The default image resolution of 640 x 640 pixels was maintained. At 100 epochs, the YOLO v8 baseline and pre-trained models produced the outcomes detailed in Table II to VII.

TABLE II
NO OF INSTANCES FOR EACH CLASS FOR YOLOV8 TRAINING

CLASS	INSTANCES
Fatigue	4
Longitudinal	30
Patch	18
Pothole	14
Transverse	21
All	87

TABLE III
RESULTS FROM YOLOV8 MODEL TRAINED AT 100 EPOCHS

	NEW MODEL		PRE-TRAINED MODEL	
CLASS	PRECISION	RECALL	PRECISION	RECALL
Fatigue	0.000	0.000	1.000	0.250
Longitudinal	1.000	0.033	0.419	0.167
Patch	0.769	0.556	0.652	0.833
Pothole	0.500	0.143	0.500	0.286
Transverse	1.000	0.048	1.000	0.048
All	0.654	0.156	0.714	0.317

TABLE IV
RESULTS FROM YOLOV8 MODEL TRAINED AT 200 EPOCHS

	NEW MODEL		PRE-TRAINED MODEL	
CLASS	PRECISION	RECALL	PRECISION	RECALL
Longitudinal	1.000	0.133	0.667	0.133
Patch	0.778	0.778	0.923	0.667
Pothole	0.800	0.286	0.833	0.357
Transverse	0.727	0.381	0.800	0.190
All	0.661	0.316	0.845	0.320

TABLE V
RESULTS FROM YOLOV8 MODEL TRAINED AT 300 EPOCHS

	NEW MODEL		PRE-TRAINED MODEL	
CLASS	PRECISION	RECALL	PRECISION	RECALL
Longitudinal	1.000	0.033	0.643	0.300
Patch	0.817	0.722	0.824	0.778
Pothole	0.833	0.357	0.800	0.571
Transverse	1.000	0.095	0.600	0.429
All	0.730	0.242	0.773	0.466

TABLE VI
RESULTS FROM YOLOV8 MODEL TRAINED AT 400 EPOCHS

	NEW MODEL		PRE-TRAINED MODEL	
CLASS	PRECISION	RECALL	PRECISION	RECALL
Longitudinal	0.625	0.167	0.667	0.267
Patch	0.824	0.778	0.833	0.556
Pothole	0.800	0.286	0.833	0.357
Transverse	0.700	0.333	0.625	0.238
All	0.601	0.341	0.800	0.362

TABLE VII
RESULTS FROM YOLOV8 MODEL TRAINED AT 500 EPOCHS

	NEW MODEL		PRE-TRAINED MODEL	
CLASS	PRECISION	RECALL	PRECISION	RECALL
Longitudinal	0.625	0.167	0.667	0.267
Patch	0.824	0.778	0.833	0.556
Pothole	0.857	0.429	0.875	0.500
Transverse	0.700	0.333	0.625	0.238
All	0.601	0.341	0.800	0.362

200 epochs produced the greatest performance for the YOLO baseline model (73%), while 400 epochs produced the best performance for the pre-trained model (84.5%). The model demonstrates a universal recall-precision trade-off, as precision scores generally decreased as recall scores increased. This indicates that the model accurately predicts instances for each category (few false positives) but ignores a significant number of actual positive instances (high false negatives).

A. Comparison to Baseline Method

To evaluate the efficacy of the proposed method for pavement distress detection, both the proposed model and YOLO v8 approaches were compared comprehensively to baseline models. The dataset of 680 pavement images captured under various lighting conditions and distress categories was used for comparison. Each image was painstakingly annotated by domain specialists to provide ground-truth labels for regions of distress. Evaluation metrics included precision, recall, and macro-weighted precision and accuracy. The proposed technique significantly outperformed the baseline method for both metrics. It attained a precision of 90.40 percent compared to the baseline's 20 percent, indicating a greater accuracy in identifying positive results. The recall rate also increased significantly, with the proposed method attaining 90.1% as op- posed to the baseline's 24.7%. As a result, the overall accuracy improved, with the proposed method yielding an impressive 89.9% as opposed to the baseline's 24.75%. These results clearly demonstrate that the proposed method is superior to the baseline technique for detecting pavement distresses. Using deep learning techniques, in particular the employed CNN architecture, enables the model to effectively capture complex patterns and variations in distress categories, resulting in enhanced detection performance.

TABLE VII
Final Comparison

BASELINE CNN		BASELINE YOLO V8		PRE-TRAINED YOLO V8		**PROPOSED MODEL**	
Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
20.00%	24.7%	73.0%	24.2%	84.5%	32.0%	**90.4%**	**90.1%**

V. Conclusion and Future Work

In conclusion, this paper has investigated and presented a comprehensive analysis of pavement distress detection using deep learning models. The experiments conducted with the proposed model architectures showcased the efficacy and potential of these models in accurately identifying and classifying various types of distresses. The method outperformed the base- line approach with a macro-averaged precision score of 90.4% and macro-averaged recall score of 90.1%, demonstrating superior performance. It also outperformed the contemporary YOLOv8 model, which yielded macro-averaged precision and recall scores of 84.5% and 32.0%, respectively.

The utilization of image processing techniques, as outlined in Figure 1, and their impact on model performance have highlighted the importance of specific techniques such as im- age grayscaling, background subtraction, and image resizing. The paper has also identified various challenges that may arise during image processing and model building.

Future endeavors will therefore focus on refining and optimizing the proposed method, exploring new architectures, additional image processing techniques, and expanding its applicability to other data formats and case scenarios. By further optimizing the data pre-processing, image processing, and classification pipeline, the efficiency of the proposed models can be improved, enabling faster and more accessible deployment. Additionally, incorporating advanced image augmentation techniques, evaluating the models on equally weighted target categories, and testing them on video data will contribute to enhancing their performance and adaptability in real-world scenarios.

In summary, this research has demonstrated the effective- ness of deep learning models in pavement distress detection and has provided insights into the importance of image processing techniques and challenges in model building. The proposed models and future directions outlined in this paper have the potential to advance the field of automated pavement condition monitoring, leading to more accurate and efficient detection of distresses and facilitating timely maintenance activities.

References

[1] David H Hubel and Torsten N Wiesel. Receptive fields of single neu- rones in the cat's striate cortex. *The Journal of physiology*, 148(3):574, 1959.

[2] Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4):193–202, 1980.

[3] Yann LeCun, Yoshua Bengio, et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995, 1995.

[4] Yao Sun, Ezzatollah Salari, and Ellie Chou. Automated pavement distress detection using advanced image processing techniques. In *2009 IEEE International Conference on Electro/Information Technology*, pages 373–377. IEEE, 2009.

[5] Stefania C Radopoulou and Ioannis Brilakis. Patch detection for pavement assessment. *Automation in Construction*, 53:95–104, 2015.

[6] Jessie R Balbin, Carlos C Hortinela IV, Ramon G Garcia, Sunnycille Baylon, Alexander Joshua Ignacio, Marco Antonio Rivera, and Jaimie Sebastian. Pattern recognition of concrete surface cracks and defects using integrated image processing algorithms. In *Second International Workshop on Pattern Recognition*, volume 10443, pages 40–44. SPIE, 2017.

[7] Jihong Liu, Jiaxin Gu, and Shan Luo. Research on road crack detection based on machine vision. In *2022 IEEE 6th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, pages 543–547. IEEE, 2022.

[8] S Nienaber, Marthinus J Booysen, and RS Kroon. Detecting potholes using simple image processing techniques and real-world footage. *South African Transport Conference*, 2015.

[9] Neelam Dwivedi, Dushyant Kumar Singh, and Dharmender Singh Kush- waha. An approach for unattended object detection through contour formation using background subtraction. *Procedia Computer Science*, 171:1979–1988, 2020.

[10] Yong Shi, Limeng Cui, Zhiquan Qi, Fan Meng, and Zhensong Chen. Automatic road crack detection using random structured forests. *IEEE Transactions on Intelligent Transportation Systems*, 17(12):3434–3445, 2016.

[11] Miguel Gavila´n, David Balcones, Oscar Marcos, David F Llorca, Miguel A Sotelo, Ignacio Parra, Manuel Ocan˜a, Pedro Aliseda, Pedro Yarza, and Alejandro Am´ırola. Adaptive road crack detection system by pavement classification. *Sensors*, 11(10):9628–9657, 2011.

[12] Nhat-Duc Hoang and Quoc-Lam Nguyen. A novel method for asphalt pavement crack classification based on image processing and machine learning. *Engineering with Computers*, 35:487–498, 2019.

[13] Dejin Zhang, Qingquan Li, Ying Chen, Min Cao, Li He, and Bailing Zhang. An efficient and reliable coarse-to-fine approach for asphalt pavement crack detection. *Image and Vision Computing*, 57:130–146, 2017.

[14] Lei Zhang, Fan Yang, Yimin Daniel Zhang, and Ying Julie Zhu. Road crack detection using deep convolutional neural network. In *2016 IEEE international conference on image processing (ICIP)*, pages 3708–3712. IEEE, 2016.

[15] Hong-Hu Chu, Muhammad Rizwan Saeed, Javed Rashid, Muhammad Tahir Mehmood, Israr Ahmad, Rao Sohail Iqbal, and Ghulam Ali. Deep learning method to detect the road cracks and potholes for smart cities. *CMC-COMPUTERS MATERIALS & CONTINUA*, 75(1):1863–1881, 2023.

[16] Emmanuel Pintelas, Ioannis E Livieris, Sotiris Kotsiantis, and Panagiotis Pintelas. A multi-view-cnn framework for deep representation learning in image classification. *Computer Vision and Image Understanding*, 232:103687, 2023.

[17] Young-Jin Cha, Wooram Choi, and Oral Bu¨yu¨ko¨ztu¨rk. Deep learning-based crack damage detection using convolutional neural networks. *Computer-Aided Civil and Infrastructure Engineering*, 32(5):361–378, 2017.

[18] Amjad Issa, Haya Samaneh, and Mohammad Ghanim. Predicting pavement condition index using artificial neural networks approach. *Ain Shams Engineering Journal*, 13(1):101490, 2022.

[19] Virender Singh, Swati Singh, and Pooja Gupta. Real-time anomaly recognition through cctv using neural networks. *Procedia Computer Science*, 173:254–263, 2020.

[20] Vishal Mandal, Lan Uong, and Yaw Adu-Gyamfi. Automated road crack detection using deep convolutional neural networks. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 5212–5215. IEEE, 2018.

[21] Ju Huyan, Wei Li, Susan Tighe, Junzhi Zhai, Zhengchao Xu, and Yao Chen. Detection of sealed and unsealed cracks with complex backgrounds using deep convolutional neural network. *Automation in Construction*, 107:102946, 2019.

[22] Liang Song and Xuancang Wang. Faster region convolutional neural network for automated pavement distress detection. *Road Materials and Pavement Design*, 22(1):23–41, 2021.

[23] Mark P Melegrito, Alvin Sarraga Alon, Sammy V Militante, Yolanda D Austria, Myriam J Polinar, and Maria Concepcion A Mirabueno. Abandoned-cart-vision: Abandoned cart detection using a deep object detection approach in a shopping parking space. In *2021 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, pages 1–5. IEEE, 2021.

[24] Jhon Michael C Manalo, Alvin Sarraga Alon, Yolanda D Austria, Nin˜o E Merencilla, Maribel A Misola, and Ricky C Sandil. A transfer learning-based system of pothole detection in roads through deep convolutional neural networks. In *2022 International Conference on Decision Aid Sciences and Applications (DASA)*, pages 1469–1473. IEEE, 2022.