Journal of Applied Research and Technology

www.jart.icat.unam.mx

Journal of Applied Research and Technology 22 (2024) 586-598

Original

Ensemble-based approach using inception V2, VGG-16, and Xception convolutional neural networks for surface cracks detection

A. Hussain[a]* • A. Aslam[b]

[a]*School of Electronics and Control Engineering, Chang'an University, Xi'an, China*
[b]*School of Information Engineering, Chang'an University, Xi'an, China*

Received 01 07 2024; accepted 06 17 2024
Available 08 31 2024

Abstract: Manual road crack detection is time-consuming. However, deep learning-based solutions are quick and accurate. Various deep learning-based convolutional neural networks (CNN) have been recently proposed. This study implies a comprehensive assessment of the performance of inception V2, VGG16, and Xception CNN utilizing the surface cracks dataset. The research approach comprises four distinct steps. Training and validating these pre-trained models are necessary by immobilizing certain foundational layers. The previously frozen layers are thawed during the second stage, and the training and validation process is repeated. Subsequently, the performance of the model is evaluated. To enhance the performance of the models in detecting surface cracks in dataset images, after completion of the model training and validation process for both frozen and unfrozen layers, the models are combined using the ensemble technique to increase the overall performance for surface crack detection. The performance of the models, including inception V2, VGG16, Xception, and the ensemble model, is evaluated using evaluation metrics including accuracy, precision, recall, and F1 score. The ensemble has the highest precision 99.97% and the highest recall 99.92%. along with the highest accuracy 99.93% and F1 score 99.92%, compared to the other CNN models.

Keywords: Road cracks, crack detection, inception V2, VGG16, Xception CNN, ensemble model

*Corresponding author.
E-mail address: 2022032907@chd.edu.cn (A. Hussain).
Peer Review under the responsibility of Universidad Nacional Autónoma de México.

1. Introduction

Roadways are an essential public resource that brings audible and durable benefits to the community. They do this because ensuring the free flow of goods and people throughout the market contributes to the growth of the market's economy. The timely maintenance of roads is critical in fostering sustainable development and competitiveness within national, local, and regional economies. Road infrastructure preservation and upkeep are vital in maximizing its advantages and ensuring long-term sustainability. To accomplish this objective, road maintenance authorities necessitate sufficient support and efficient tools to attend to matters about road conditions expeditiously (Feng et al., 2019). Insufficient road conditions lead to many problems, including hindering road navigation and increasing the probability of accidents for drivers, escalating the costs associated with vehicle maintenance for road users, and raising the expenses of repairing roads because of the road surfaces and substructures' irreversible deterioration. Numerous endeavors have been undertaken to effectively assess the condition of asphalt roads, especially road damage inspection tasks, to address the need for robust tools. In many developing countries, as well as in the past, transportation authorities have manually conducted the evaluation of pavement conditions. According to (Koch et al., 2013) and (Radopoulou & Brilakis, 2017), the method is characterized by a significant lack of efficiency, primarily due to the labor-intensive nature of conducting measurements, recording data, and processing information during field inspections. This is further compounded by the need for manual labor to conduct these tasks. Moreover, this methodology is susceptible to the potential subjectivity and biases of the technicians tasked with conducting these inspections (Guan et al., 2014). Automated techniques for identifying and categorizing road damage have garnered heightened attention from transportation agencies.

In recent years, substantial advancements have been made in automatic pavement crack identification models (Zakeri et al., 2017), largely due to image processing techniques (ITPs) advancements. The models utilized in this study perform efficient and accurate evaluations of two-dimensional (2D) road pavement images to identify and evaluate relevant features for binary crack detection. Additionally, these models appropriately label images for subsequent analyses. Although there have been advancements, the efficacy of crack detection through image processing is still limited by factors such as inadequate background lighting and the heterogeneous texture of pavement aggregate (Li et al., 2017). The challenges mentioned above have the potential to negatively affect the effectiveness of intensity-thresholding techniques, resulting in a notable rise in the occurrence of both type I and type II errors. Improving damage detection accuracy has gained much attention by utilizing local edge detection methods and general global transformation. The detection of damage contours has been a prominent area of research, with the application of various techniques such as Sobel and Canny edge detectors, Haar transform (FHT), and Fast Fourier Transform (FFT). However, edge detection techniques-based models demonstrate high efficiency and speed; their capability is restricted to a single type of damage detection and contains substantial error rates when facing distortion, lighting, and data noise challenge (Koziarski & Cyganek, 2017; Zhang et al., 2017).

Artificial neural networks (ANN), support vector machines (SVM), and classification and regression trees (CART) are some of the machine learning (ML) techniques and have been extensively employed in the field of damage recognition applications (Butcher et al., 2014; Hoang, 2018; Karmel et al., 2018; Kyriakou et al., 2019; Song et al., 2018). Image processing techniques (IPTs) are used to extract features from images; however, machine learning-based algorithms are used to determine whether they indicate a certain sort of impairment. Most research directly choosing features for extraction in information processing tasks is a bad idea that hurts the performance of machine learning models (Chow et al., 2020; Fang et al., 2020). Using models based on deep learning for road crack detection has recently been identified as an effective approach (Cha et al., 2017). Recently, much research has focused on deep learning (DL) algorithms. The algorithms can automatically and hierarchically extract representative and discriminative properties from the lowest-level feature, such as edge or texture, all the way up to the underlying feature. This aspect has been a main area of investigation in numerous studies (Fang et al., 2018; Zhang et al., 2016; Zhong et al., 2019). Subsequent systems, including faster R-CNN, R-FCN, and SSD, have significantly improved object detection accuracy. These systems are built upon single short multi-box detectors and region-based convolutional neural networks. Over time, numerous systems have been developed and enhanced. Additionally, it has been claimed that several feature extractors, including inception (Ioffe & Szegedy, 2015), residual network (Resnet) (He et al., 2016), and MobileNet (Sandler et al., 2018), work in tandem with these systems to improve detection efficiency and speed.

Artificial Intelligence, especially machine learning and deep learning, has been employed in various domains including disease prediction (Hussain & Aslam, 2024a), intrusion detection (Hussain, Khatoon, et al., 2024), fraud detection (Hussain & Hussain 2024b), facial assimilation (Hussain et al., 2023) and cracks detection (Hussain, Qureshi et al., 2024). To automate road crack detection, much research works implement deep learning-based CNN models to detect

the road cracks. The main objective of this research work is as follows:

- To implement the inception V2, VGG 16, and Xception CNN models for road crack detection using a public dataset, the models are trained using the training set. The base layers of these models are frozen during the first training process using training and validation sets. Then, the layers are unfrozen to perform the training and validation again.
- Furthermore, after completing the training and validation process, the three models are combined using the ensemble technique to enhance their performance.
- The performance evaluation of the ensemble-based and three-based models uses several performance metrics, including accuracy, precision, recall, and F1 score.

This paper is organized as follows: Section II overviews well-known CNN models. Section III contains the methodology and dataset, while Section IV contains experimentation of three models, training and validation results on freeze and unfreeze layers, and the ensemble process along with the cracks detection. Section V contains the evaluation metrics and the model's performance, including the confusion matrix of the models and the performance comparison. Section VI provides the conclusion of this study.

2. Related work

Several studies have proposed using automated crack detection techniques as a substitute for manual inspections. A substantial body of literature has emerged in recent decades concerning detecting cracks in structural surfaces, like roads, bridges, pavements, and tunnel walls. Research studies evaluating the works can be accessed (Attard et al., 2018; Koch et al., 2015; Wang & Huang, 2010). Many image processing techniques have been used. Previous work relied on various approaches, including mathematical morphology, thresholding, and edge detection. Various new techniques have been utilized to explore the identification of cracks in challenging environments. The methodologies involve applying different alternative techniques, including machine learning, wavelet transform, texture analysis, saliency detection, and minimal pathfinding. Although these techniques have demonstrated their utility in various scenarios, they are constrained by the drawbacks inherent in rule-based methodologies and superficial abstractions when confronted with images of cracks. The elements include cracks' inhomogeneity, surface texture variability, background complexity, crack noise identification (such as joints), and crack topology's inherent difficulty. The challenges mentioned above make it impractical to use a rule-based approach, as it may not efficiently extract general features when confronted with changing circumstances. Researchers have proposed a new deep learning method to overcome these challenges, specifically employing convolutional neural networks (CNNs). This approach provides a higher level of abstraction and generalization, eliminating the need for manually engineered feature extraction.

Alex (2012) proposed 2012 an AlexNet network structure consisting of three fully connected layers and five convolution layers (convolution + nonlinear activation + maximum pooling layer). The Rectified Linear Unit (ReLU) was employed to address gradient divergence within the network. At the same time, the dropout technique was implemented in the fully connected layer to mitigate the overfitting of the network. In 2014, VGGNet (Simonyan & Zisserman, 2014), which had models of networks with depths of 11 to 19 layers, came out. The most popular types were the VGG16 and VGG19, containing five convolution layers, including three SoftMax output layers and three full link layers. The results of the model showed that raising the depth of detection made the accuracy better. Also, GoogleNet (Szegedy et al., 2015) introduced the "network in network" based inception architecture (Liu et al., 2021), which replaced the optimal local coefficient structure with dense components. The training time and generalization capacity improved by replacing more parameters with a high number of 11 convolution kernels within the inception structure. Two classifiers were added to the model to help with the gradient propagation and alleviate the problem of vanishing gradients. A pair of 33 convolution kernels was used instead of the original 55 convolution kernel in the inception-v2 model (Ioffe & Szegedy, 2015). It used a convolution integral approach to get the same result with fewer parameters and faster computation. The inception-v3 (Szegedy et al., 2016) model introduced the concept of decomposition, which involves breaking down the n x n convolution kernel into separate n x 1 and 1 x n convolution kernels. This technique aims to enhance the depth and nonlinearity of the network.

In 2015, ResNet (He et al., 2014) came up with the idea of skipping the raw data to go straight to the output. Also, the difference between the input and the output, the residual value, was added as a change to the direct learning goal value. Installing the skip connection design has made the learning goal less complicated and helped solve some knowledge loss and degradation problems. In 2016, Szegedy et al. (2017), proposed inception-Resnet-v2 and inception-v4. The inception-Resnet-v2 was built around the inception-v4 framework but with increased precision. Residual connections were used instead of filter concatenation in the inception-Resnet-v2 model, which improved training and performance. DenseNet (Huang et al., 2017) achieved inter-layer connectivity in the year of its launch, efficiently utilizing features and addressing the problem of disappearing gradients. The training effectiveness showed a considerable improvement over Resnet. Chollet F unveiled the Xception

convolutional neural network architecture in 2017 (Chollet, 2017). This design uses a residual connection method that speeds up convergence and increases accuracy. It is based on the profoundly separable convolution layer.

2.1. Discussion

Several researchers have performed surface crack detection. However, most of the existing work implements these CNNs individually. Also, most researchers do not implement layer freezing while training the models, whereas freezing the layers of the CNN models using training can enhance the model performance for detection, i.e., road crack detection. Meanwhile, the model's training and validation can be performed again by unfreezing the base layers. Furthermore, combining these models together using an ensemble technique can enhance the model's overall performance. This research compares the models' performances during training and validation using the freezing and unfreezing layers. The crack detection is performed for the pre-trained models, along with the ensemble-based model, after the training and validation process, which consists of two phases, where the base layers are frozen during the first training and validation process, and the layers are then unfrozen to perform the training process again.

3. Methodology

We are using three well-known pre-trained CNN, inception V2, VGG16, and Xception CNN model for crack detection using a public dataset surface cracks dataset, available publicly on Kaggle, containing 40000 images of cracks(positive) and non-cracks(negative). Each CNN model is used for training and validation in two phases. In the first phase, the base layers are frozen during training and validation, with the phase output shown in graphical form. For the second phase, the layers are unfrozen, and training and validation are performed again, where the output results are also shown in graphical form. The accuracy score for each CNN model is generated later. During the test phase, the input images contain cracks and non-cracks, whereas the output images highlight the image features like cracks, etc., using a heat map. The confusion matrix for each CNN model is generated for the performance evaluation. An ensemble is performed to increase the test output, where all three CNN models are combined for crack detection; the ensemble's output shows improvement as all three models are now combined for detecting cracks. The methodology of this study is shown in Figure 1.

3.1. Dataset

The Surface crack dataset contains two groups for image classification: Without cracks (negative) and with cracks (positive). This dataset includes 40,000 images of 227 x 227 pixels with RGB channels. There are 20,000 photos in each class. Zhang et al. (2016) applied the method to 458 high-resolution images with a 4032x3024 pixel dimension to create this dataset. Significant variations in surface quality and illumination were visible in the high-resolution photographs. No random flipping, tilting, rotating, or other types of data augmentation were used. Figures 2 and 3 below show examples of positive (cracks) and negative (no cracks) situations.

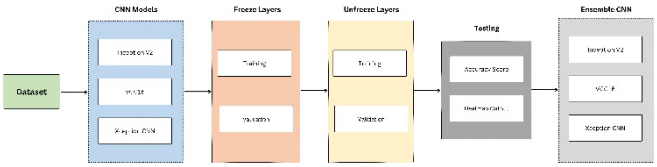

Figure 1. Research methodology.

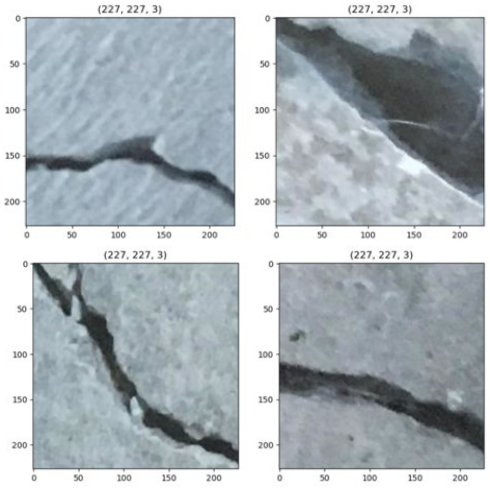

Figure 2. Positive images – with cracks.

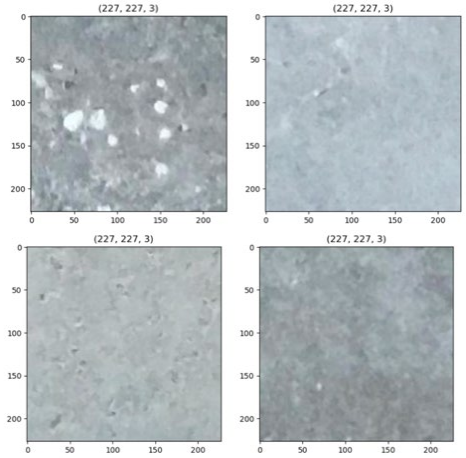

Figure 3. Negative images – without cracks.

4. Implementation

The CNN models are implemented using the surface cracks dataset having 40000 images of cracks(positive) and non-cracks(negative) with the data split for training, validation, and testing using 60%, 20%, and 20%, respectively. The training of each CNN is done by freezing and unfreezing some of the layers of the CNN model for better training; the layers are frozen for the inception V2, VGG 16, and Xception CNN models in the first phase. In the second phase, the layers are unfrozen. Training and validation are performed again, and in the end, the accuracy score of each model is generated, along with the output of the images using a heat map with the GridCam. The heat map highlights the image features, especially the cracks. The performance of CNN models during testing and validation in freeze and unfreeze layers are shown in graphical form to compare the results at each phase. The confusion matrix for the testing outputs is generated to compare the performance of the CNN models for the dataset used.

4.1. Inception V2

The inception-ResNet-v2 CNN has undergone training using a dataset of over one million images from the ImageNet database. The neural network architecture consists of 164 layers, enabling it to accurately classify images into a wide range of object categories. These categories include a wide variety of things, such as keyboards, mice, pencils, and many different animals, but they are not restricted to just those things. That is why the neural network has obtained extensive feature representations for many images. The image input size of the network is 299 x 299 pixels. Inception v2 is the next version of inception convolutional neural network architectures. It is distinguished by the inclusion of batch normalization as a prominent feature. Additional modifications encompass eliminating dropout and excluding local response normalization, owing to the advantageous effects of batch normalization.

4.1.1. Freezing layers

Firstly, the layers are frozen, and training and validation are performed. The validation results for accuracy and loss are better than the training results; Figures 4 and 5 show the training and validation accuracy and loss.

4.1.2. Unfreezing layers

The layers are unfrozen after training and validation on freeze layers, and training and validation are performed again. This time, the validation accuracy and loss are better than the training accuracy and loss. Figure 6 and Figure 7 show training and validation accuracy and training and validation loss.

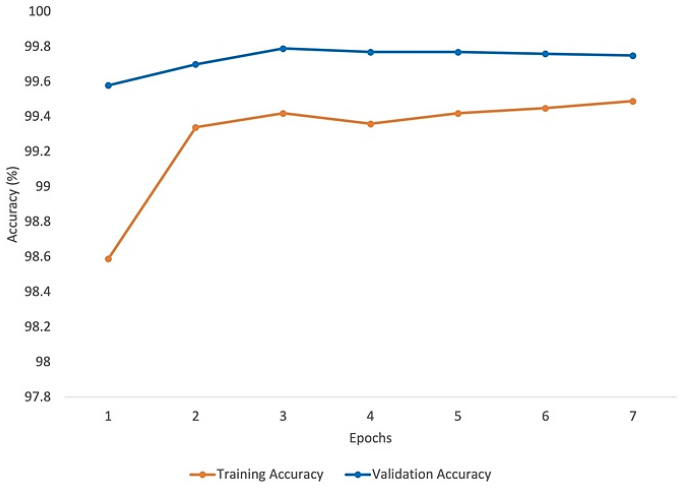

Figure 4. Inception V2 – training and validation accuracy on freeze layers.

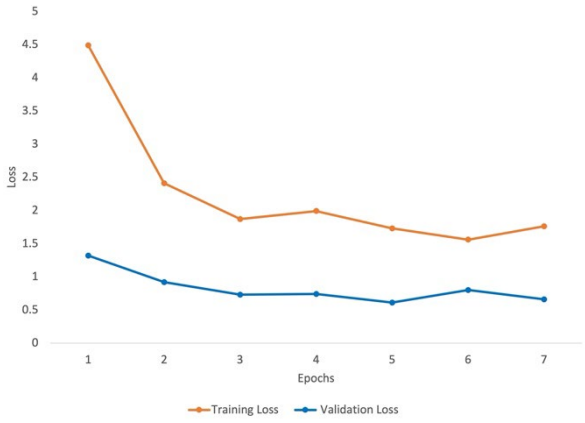

Figure 5. Inception V2 – training and validation loss on freeze layers.

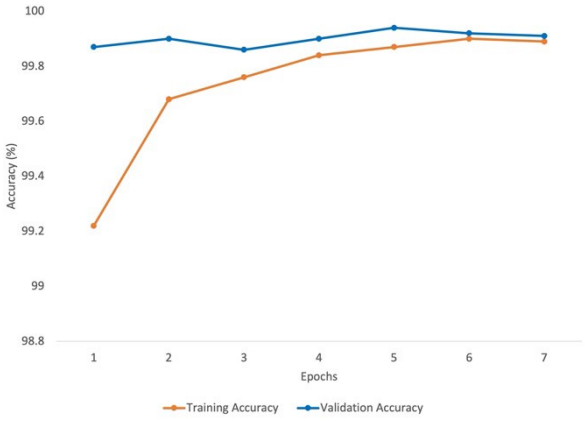

Figure 6. Inception V2 – training and validation accuracy on unfreeze layers.

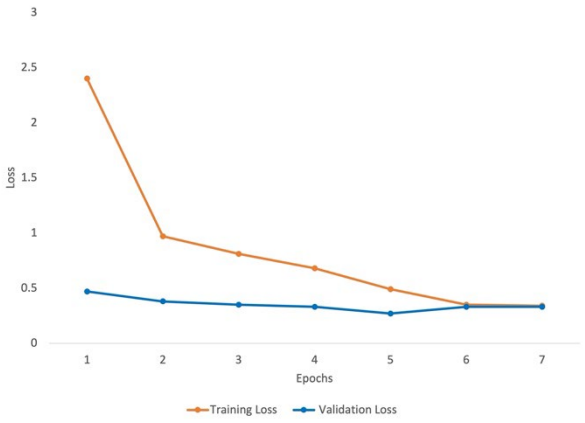

Figure 7. Inception V2 – training and validation loss on unfreeze layers.

The accuracy score for inception V2 has also generated images using testing, where the output of the crack and non-crack images is generated by using a heat map to highlight the image features, as shown in Figure 8.

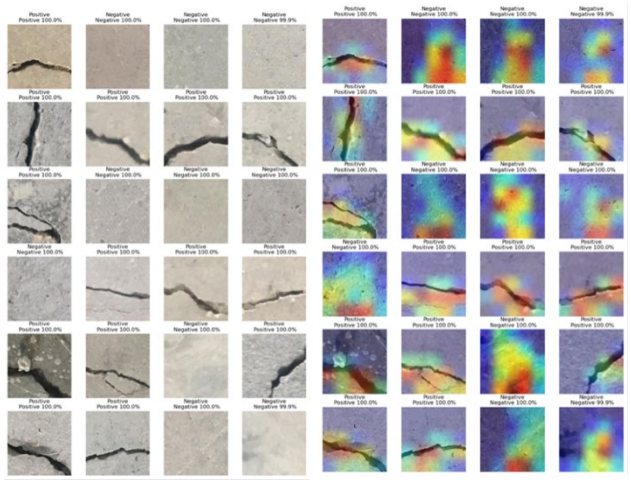

Figure 8. Inception V2 – input images and output images with heat map.

4.2. VGG16

The VGG-16 CNN model has 16 layers with an image input size of 224 x 224 pixels. It is feasible to use a neural network that has already been trained using data from the ImageNet database, which contains more than a million photos. The existing neural network can effectively classify images into a wide range of 1000 object categories, including keyboards, mice, pencils, and numerous animal species. So, the model has successfully obtained extensive feature representations for a wide range of images.

4.2.1. Freezing layers

To enhance the VGG16's performance for crack detection, we utilize the VGG16 base model and freeze part of the layers during the first phase of training and validation; the results reveal that the validation accuracy and loss are marginally better than the training accuracy and loss. Figures 9 and 10 display the outcomes of the training and validation processes for the freeze layers.

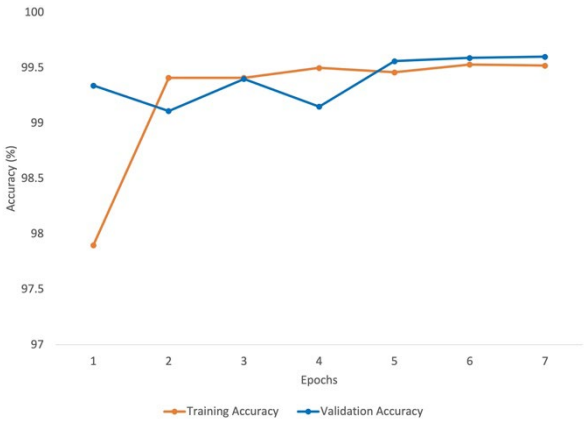

Figure 9. VGG16 – training and validation accuracy on freeze layers.

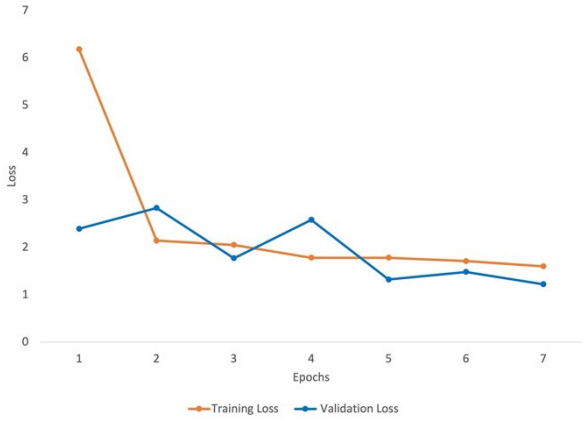

Figure 10. VGG16 – training and validation loss on freeze layers.

4.2.2. Unfreezing layers

After training and validation on freeze layers, the layers of the base model are unfrozen. The training and validation are performed again to improve the model performance. The Training and the validation results for unfreezing layers in Figure 11 and Figure 12 indicate that the validation accuracy surpasses the training accuracy. Additionally, the validation loss is lower than the training loss.

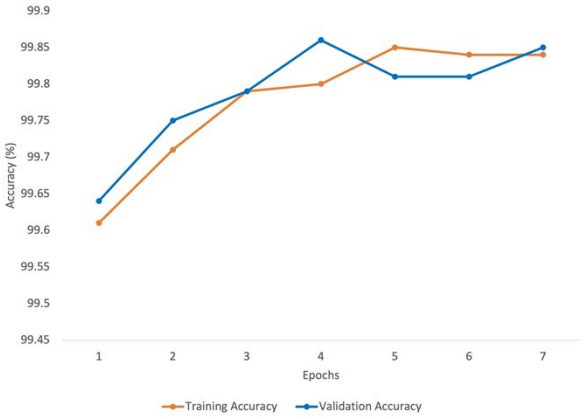

Figure 11. VGG16 – training accuracy and validation accuracy on unfreeze layers.

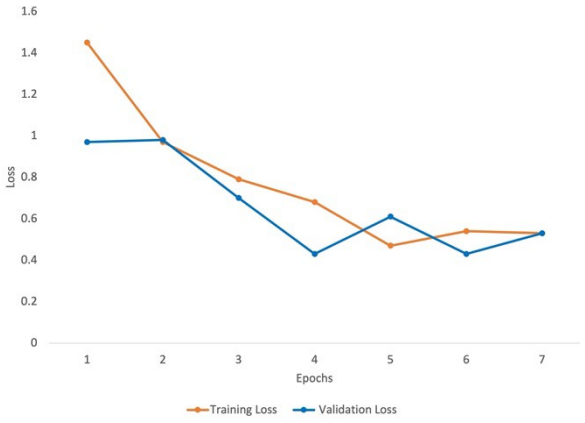

Figure 12. VGG16 – training accuracy and validation loss on unfreeze layers.

After performing training and validation in the first two phases, the accuracy score of VGG16 is generated along with the output of the crack and non-crack images using testing. A heat map is used to highlight the image features, as shown in Figure 13.

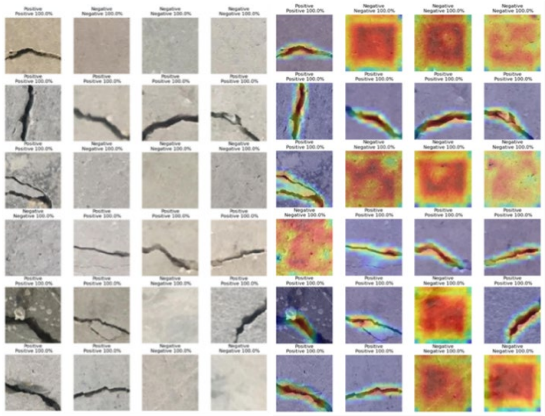

Figure 13. VGG16 – input images and output images with heat map.

4.3. Xception

The Xception CNN architecture is known for its significant depth, comprising 71 layers with an image input size of 299 x 299 pixels. The neural network has been trained using more than one million images from the ImageNet database. A pre-trained neural network model can accurately classify images into a wide range of 1000 different object categories, including keyboards, mice, pencils, and a diverse range of animal species. As a result, the neural network has successfully obtained extensive feature representations for a wide range of images.

4.3.1. Freezing layers

In the first phase, the Xception CNN model is used for training and validation, and some of the base layers are frozen, which will help improve the model's performance. The validation accuracy shows better results than the training accuracy. Additionally, the validation loss demonstrates a lower value than the training loss, as depicted in Figures 14 and 15.

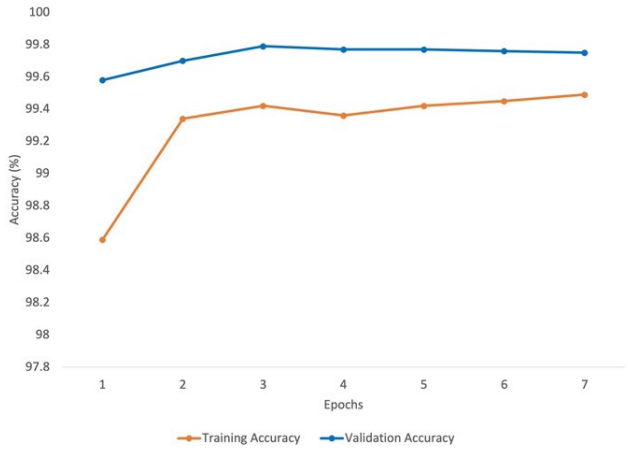

Figure 14. Xception – training accuracy and validation accuracy on freeze layers.

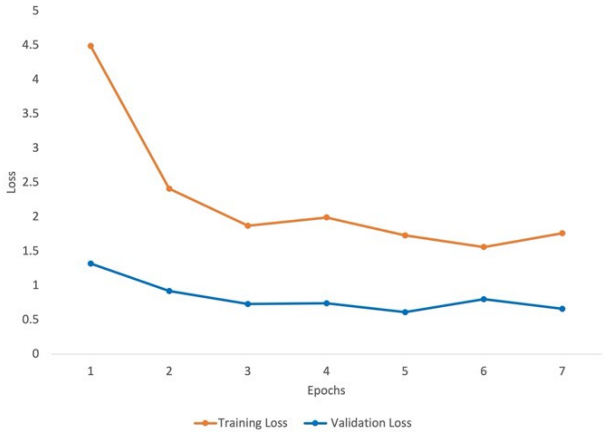

Figure 15. Xception – training accuracy and validation loss on freeze layers.

4.3.2. Unfreezing layers

The base layers of the Xception CNN model are unfrozen in the second phase of training and validation. Figures 16 and 17 demonstrate that the validation performance, as indicated by the validation accuracy and validation loss, surpasses that of the training accuracy and loss.

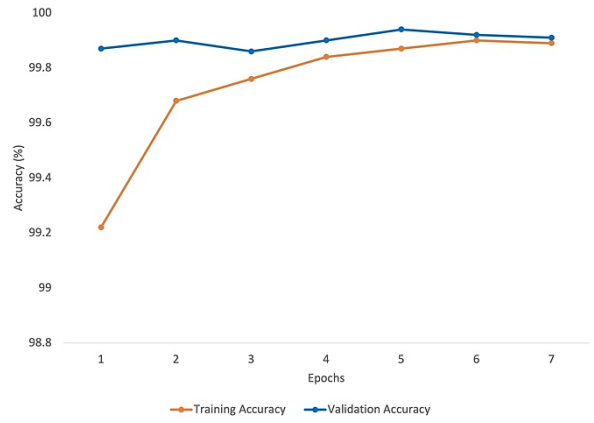

Figure 16. Xception – training accuracy and validation accuracy on unfreeze layers.

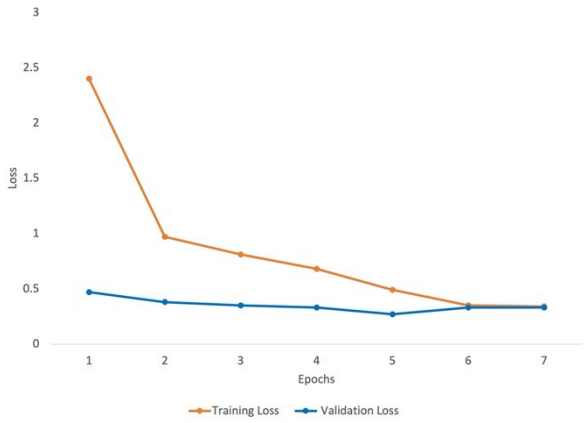

Figure 17. Xception – training accuracy and validation loss on unfreeze layers.

The accuracy score of the Xception CNN model is generated, the testing is performed, and the output images show the image features highlighted because of the use of a heat map, especially the image cracks, as shown in Figure 18.

4.4. Ensemble model

Ensemble learning, called "ensemble-based model," is a widely recognized collection of machine learning and statistical techniques that enhance prediction performance through diverse learning algorithms. The ensemble's predictions demonstrate higher accuracy when compared to the predictions made by any individual model within the ensemble. Ensemble methods consist of a wide array of approaches that exhibit varying levels of complexity. Our current focus is on aggregating forecasts produced by multiple pre-trained deep-learning networks. Different networks demonstrate unique errors, and the ensemble method can be utilized to harness the collective impact of these errors. Applying ensemble-based models in deep learning has shown impressive results, although it is not as extensively covered in the deep learning literature as traditional machine learning methods. The approach utilized in this investigation involved the integration of three models for training and validation purposes. The ensemble model is executed without utilizing the heat map to generate output images. The output sample of the ensemble is shown in Figure 19.

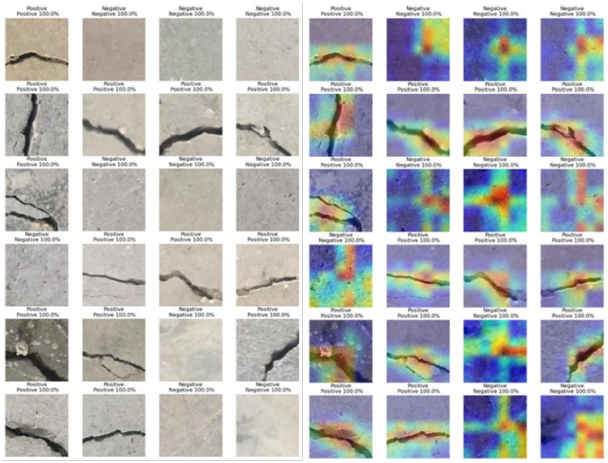

Figure 18. Xception – input images and output images with heat map.

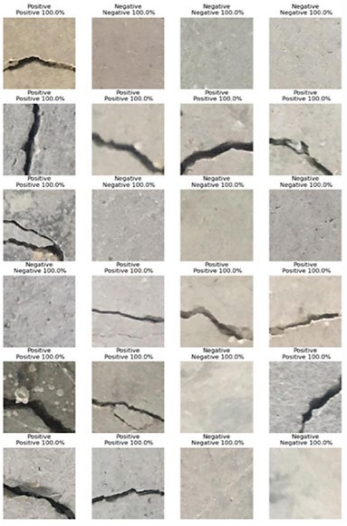

Figure 19. Ensemble process output images.

5. Results and discussions

5.1. Evaluation metrics

Various evaluation criteria, such as accuracy, precision, recall, and the F1 score, were used to compare the experimental outcomes fairly. Accuracy is a quantitative metric that can be computed by evaluating the ratio of correctly identified crack and non-crack patches to the total number of input patches. This calculation is represented by Equation 1.

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + TN + FP} \quad (1)$$

True positive (TP) and true negative (TN) refer to correctly identifying crack and non-crack patches. In contrast, FP (false positive) and FN (false negative) refer to erroneously identifying crack and non-crack patches. According to Equation 2, precision is the fraction of correctly detected crack patches relative to the total number of crack patches identified by the classifier.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

As indicated in Equation 3, recall can be defined as the ratio of correctly detected crack patches to the total number of crack patches.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

The F1 score is computed by taking the average of the model's recall and precision using the same calculation method. The F1 score is mathematically represented by Equation 4.

$$\text{F1 score} = 2 \text{ x } \frac{\text{Precision x Recall}}{\text{Precision + Recall}} \quad (4)$$

5.2. Confusion matrix

The confusion matrix is generated to evaluate the inception V3, VGG16, and Xception CNN. This matrix provides an overview of each model's overall performance. The dataset comprises 40,000 images, consisting of both crack and non-crack images. However, only 8,000 images are utilized for testing purposes across the three models. The confusion matrix for the inception V2, VGG16, and Xception CNN is shown in Figures 20, 21, and 22.

The output of the ensemble process is shown in the confusion matrix, which highlights the performance of the ensemble of three models and shows a slight improvement. The performance of the ensemble process is shown in Figure 23.

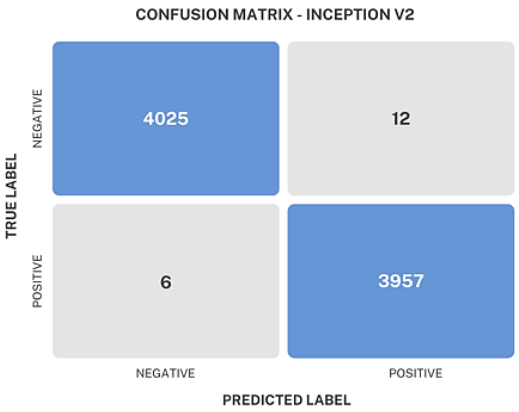

Figure 20. Confusion matrix of inception V2 model.

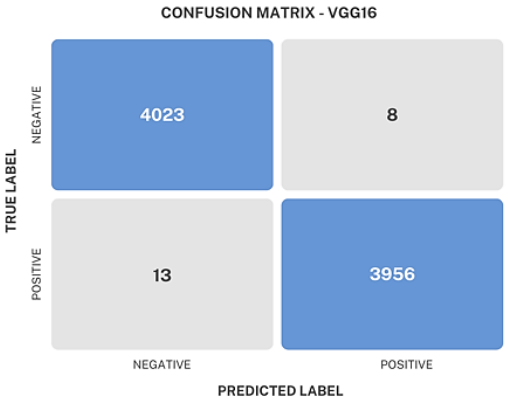

Figure 21. Confusion matrix of VGG16 model.

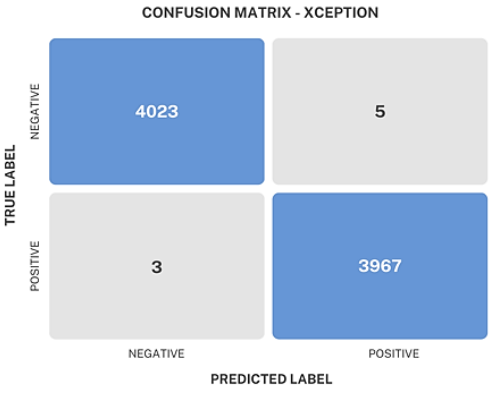

Figure 22. Confusion matrix of Xception model.

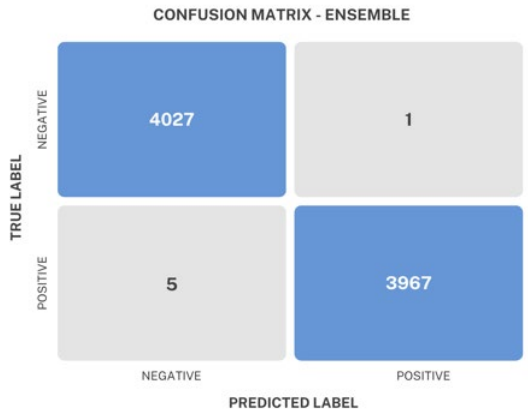

Figure 23. Confusion matrix of ensemble-based model.

5.3. Performance evaluation

The comparison table highlights the efficacy of an ensemble approach in machine learning, notably for detecting surface cracks. When we compare specific models such as inception V2, VGG-16, and Xception, we see satisfactory performance across all metrics—precision, recall, accuracy, and F1 score—with Xception outperforming the others. However, the ensemble method, which combines the capabilities of these separate models, outperforms them, achieving the highest precision of 99.97%, the highest recall of 99.87%, and the highest accuracy of 99.93%. The ensemble also has the greatest F1 score, which balances precision and recall, at 99.92%. This demonstrates the ensemble's superior capacity to locate cracks while correctly retaining a low false positive rate, highlighting the ensemble model's robustness and reliability in real applications. The comparison of the model's performance is shown in Table 1.

Table 1. Performance comparison.

Models	Precision	Recall	Accuracy	F1 score
Inception V2	99.70%	99.85%	99.78%	99.77%
VGG 16	99.80%	99.67%	99.74%	99.74%
Xception	99.87%	99.92%	99.90%	99.90%
Ensemble model	99.97%	99.87%	99.93%	99.92%

The test performance of the models is also shown in Figure 24 below.

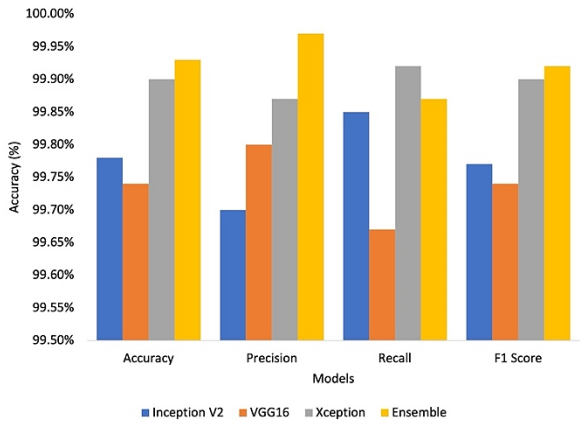

Figure 24. Test performance comparison.

6. Conclusions

Road maintenance organizations can efficiently repair road surfaces, maintain optimal road conditions, optimize transportation safety, and save transportation costs by identifying road problems quickly and accurately. Multiple convolutional neural networks (CNN) have recently been suggested. The performance of inception V2, VGG16, and Xception CNN was evaluated using the surface cracks dataset. The research methodology comprises four sequential steps. The first phase involves freezing some base layers and training and validating these pre-trained models. In the second step, the layers are unfrozen, and the process of training and validation is repeated; then, the models are evaluated using the ensemble technique, and these three models are merged to enhance their performance for surface crack detection dataset images. All four models perform well, including inception V2, VGG16, Xception, and the ensemble model. Xception has the highest recall (99.92%), while ensemble has the most precision (99.97%). However, the ensemble model has the best overall balance, with the highest accuracy (99.93%) and F1 score (99.92%). This proves that integrating multiple models to improve predictive performance is highly effective.

Conflict of interest

The authors have no conflict of interest to declare.

Funding

The authors received no specific funding for this work.

References

Attard, L., Debono, C. J., Valentino, G., & Di Castro, M. (2018). Tunnel inspection using photogrammetric techniques and image processing: A review. ISPRS journal of photogrammetry and remote sensing, 144, 180-188. https://doi.org/10.1016/j.isprsjprs.2018.07.010

Butcher, J. B., Day, C. R., Austin, J. C., Haycock, P. W., Verstraeten, D., & Schrauwen, B. (2014). Defect detection in reinforced concrete using random neural architectures. *Computer-Aided Civil and Infrastructure Engineering*, *29*(3), 191-207. https://doi.org/10.1111/mice.12039

Cha, Y. J., Choi, W., & Büyüköztürk, O. (2017). Deep learning-based crack damage detection using convolutional neural networks. *Computer-Aided Civil and Infrastructure Engineering*, *32*(5), 361-378. https://doi.org/10.1111/mice.12263

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1251-1258).

Chow, J. K., Su, Z., Wu, J., Tan, P. S., Mao, X., & Wang, Y. H. (2020). Anomaly detection of defects on concrete structures with the convolutional autoencoder. *Advanced Engineering Informatics*, *45*, 101105. https://doi.org/10.1016/j.aei.2020.101105

Fang, Q., Li, H., Luo, X., Ding, L., Rose, T. M., An, W., & Yu, Y. (2018). A deep learning-based method for detecting non-certified work on construction sites. *Advanced Engineering Informatics*, *35*, 56-68. https://doi.org/10.1016/j.aei.2018.01.001

Fang, W., Luo, H., Xu, S., Love, P. E., Lu, Z., & Ye, C. (2020). Automated text classification of near-misses from safety reports: An improved deep learning approach. *Advanced Engineering Informatics*, *44*, 101060. https://doi.org/10.1016/j.aei.2020.101060

Feng, C., Zhang, H., Wang, S., Li, Y., Wang, H., & Yan, F. (2019). Structural damage detection using deep convolutional neural network and transfer learning. *KSCE Journal of Civil Engineering*, *23*, 4493-4502. https://doi.org/10.1007/s12205-019-0437-z

Guan, H., Li, J., Yu, Y., Chapman, M., Wang, H., Wang, C., & Zhai, R. (2014). Iterative tensor voting for pavement crack extraction using mobile laser scanning data. *IEEE Transactions on Geoscience and Remote Sensing*, *53*(3), 1527-1537. https://doi.org/10.1109/TGRS.2014.2344714

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

He, Q., Li, N., Luo, W. J., & Shi, Z. Z. (2014). A survey of machine learning algorithms for big data. *Pattern Recognition and Artificial Intelligence*, *27*(4), 327-336.

Hoang, N. D. (2018). An Artificial Intelligence Method for Asphalt Pavement Pothole Detection Using Least Squares Support Vector Machine and Neural Network with Steerable Filter-Based Feature Extraction. *Advances in Civil Engineering*, *2018*(1), 7419058. https://doi.org/10.1155/2018/7419058

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).

Hussain, A., & Aslam, A. (2024a). Cardiovascular disease prediction using risk factors: A comparative performance analysis of machine learning models. *Journal on Artificial Intelligence*, *6(1)*, 129-152. https://doi.org/10.32604/jai.2024.050277

Hussain, A. & Aslam, A. (2024b). A performance analysis of machine learning techniques for credit card fraud detection. *Journal on Artificial Intelligence*, *6(1)*, 1-21. https://doi.org/10.32604/jai.2024.047226

Hussain, A., Khatoon, A., Aslam, A., Tariq, & Khosa, M.A. (2024). A comparative performance analysis of machine learning models for intrusion detection classification. *Journal of Cyber Security*, *6(1)*, 1-23. https://doi.org/10.32604/jcs.2023.046915

Hussain, A., Ullah, A., Aslam, A., & Khatoon, A. (2023). A Modified Siamese Network for Facial Assimilation. *WSEAS Transactions on Signal Processing*, vol. 19, pp. 60-66. https://doi.org/10.37394/232014.2023.19.7

Hussain, A., Qureshi, K. N., Anwar, R. W., & Aslam, A. (2024). A Novel SCD11 CNN Model Performance Evaluation with Inception V3, VGG16 and ResNet50 Using Surface Crack Dataset. In *2024 2nd International Conference on Unmanned Vehicle Systems-Oman (UVS)* (pp. 1-7). IEEE. https://doi.org/10.1109/UVS59630.2024.10467149

Ioffe, S. & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. Proceedings of the 32nd International Conference on Machine Learning, in Proceedings of Machine Learning Research, 37, 448-456. Available from https://proceedings.mlr.press/v37/ioffe15.html

Karmel, A., Adhithiyan, M., & Kumar, P. S. (2018). Machine learning based approach for pothole detection. *International Journal of Civil Engineering and Technology (IJCIET)*, *9*(5), 882-888.

Koch, C., Georgieva, K., Kasireddy, V., Akinci, B., & Fieguth, P. (2015). A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. *Advanced engineering informatics*, *29*(2), 196-210. https://doi.org/10.1016/j.aei.2015.01.008

Koch, C., Jog, G. M., & Brilakis, I. (2013). Automated pothole distress assessment using asphalt pavement video data. *Journal of Computing in Civil Engineering*, *27*(4), 370-378. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000232

Koziarski, M., & Cyganek, B. (2017). Image recognition with deep neural networks in presence of noise–dealing with and taking advantage of distortions. *Integrated Computer-Aided Engineering*, *24*(4), 337-349. https://doi.org/10.3233/ICA-170551

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). AlexNet. *Adv. Neural Inf. Process. Syst*, 1-9.

Kyriakou, C., Christodoulou, S. E., & Dimitriou, L. (2019). Smartphone-based pothole detection utilizing artificial neural networks. *Journal of Infrastructure Systems*, *25*(3), 04019019. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000489

Li, S., Cao, Y., & Cai, H. (2017). Automatic pavement-crack detection and segmentation based on steerable matched filtering and an active contour model. *Journal of Computing in Civil Engineering*, *31*(5), 04017045. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000695

Liu, J., Li, C., Liang, F., Lin, C., Sun, M., Yan, J., ... & Xu, D. (2021). Inception convolution with efficient dilation search. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11486-11495).

Radopoulou, S. C., & Brilakis, I. (2017). Automated detection of multiple pavement defects. *Journal of Computing in Civil Engineering*, *31*(2), 04016057. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000623

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4510-4520).

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. https://doi.org/10.48550/arXiv.1409.1556

Song, H., Baek, K., & Byun, Y. (2018). Pothole detection using machine learning. *Advanced Science and Technology*, 151-155.

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 31, No. 1). https://doi.org/10.1609/aaai.v31i1.11231

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2818-2826).

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).

Wang, P., & Huang, H. (2010). Comparison analysis on present image-based crack detection methods in concrete structures. In *2010 3rd international congress on image and signal processing* (Vol. 5, pp. 2530-2533). IEEE. https://doi.org/10.1109/CISP.2010.5647496

Zakeri, H., Nejad, F. M., & Fahimifar, A. (2017). Image based techniques for crack detection, classification and quantification in asphalt pavement: a review. *Archives of Computational Methods in Engineering*, *24*, 935-977. https://doi.org/10.1007/s11831-016-9194-z

Zhang, A., Wang, K. C., Li, B., Yang, E., Dai, X., Peng, Y., ... & Chen, C. (2017). Automated pixel-level pavement crack detection on 3D asphalt surfaces using a deep-learning network. *Computer-Aided Civil and Infrastructure Engineering*, *32*(10), 805-819. https://doi.org/10.1111/mice.12297

Zhang, L., Zhang, L., & Du, B. (2016). Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and remote sensing magazine*, *4*(2), 22-40. https://doi.org/10.1109/MGRS.2016.2540798

Zhong, B., Xing, X., Love, P., Wang, X., & Luo, H. (2019). Convolutional neural network: Deep learning-based classification of building quality problems. *Advanced Engineering Informatics*, *40*, 46-57. https://doi.org/10.1016/j.aei.2019.02.009