Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2058

Research Article

Real-time pothole and crack detection for safer roads: A VGG-16 and CNN approach

Neha TANWAR[1], Anil V. TURUKMANE[1,*]

[1]*School of Computer Science Engineering, VIT-AP University, Beside AP Secretariat, Amaravati, Andhra Pradesh, 522241, India*

ARTICLE INFO

Article history
Received: 10 December 2024
Revised: 12 March 2025
Accepted: 05 May 2025

Keywords:
Automatic Detection; CNN; Road Cracks; Road Potholes; Transfer Learning; VGG-16

ABSTRACT

The detection xof surface anomalies, such as potholes and cracks, in a timely and accurate manner is crucial for road safety. In this paper, we have proposed a deep learning based approach. We have used a pre-trained VGG-16 model for robust feature extraction and a custom CNN for classification. Our model has an accuracy of 93% for detecting road anomalies and shows better performance than conventional methods. The proposed system has promising potential to be integrated into the navigation of autonomous vehicles and infrastructure of smart city in real-time analysis. In this study, a system based on a Convolutional Neural Network (CNN) has been developed to detect and classify road anomalies in real-time. The system identifies cracks, potholes, and normal road surfaces, while supporting safer autonomous driving. CNN model was trained using transfer learning. The VGG-16 model was further trained for road anomaly classification. The model was trained using a dataset of images with three types of road surface viz. cracks, potholes and normal surfaces. The model achieved an accuracy rate of 93.0% in distinguishing among these road conditions. Performance metrics were evaluated to confirm the effectiveness of the model. Matthew's correlation coefficient (MCC) was found to be 0.8934, while Cohen's kappa (CK) was 0.896. The mean square error (MSE) was calculated to be 0.0315 and the peak signal-to-noise ratio (PSNR) was 63.147 dB. Multiple evaluation criteria were used to validate the performance of the model. The obtained results were compared with various deep learning algorithms. Transfer learning was found to be an effective and reliable approach for automating road anomaly detection. This method provides a cost-effective way to make self-driving cars safer by allowing for the accurate detection of road surface anomalies in real-time.

Cite this article as: Tanwar N, Turukmane AV. Real-time pothole and crack detection for safer roads: A VGG-16 and cnn approach. Sigma J Eng Nat Sci 2026;44(3):1700−1713.

INTRODUCTION

Autonomous Vehicles Future is Revolutionary in Transportation. The integration of sophisticated systems of traffic intelligence has contributed significantly to driving safety and comfort [1]. Also, self-driving and autonomous vehicles can improve the efficiency of driving, reduce the traffic congestion and lower the emissions of pollutants [2].

*__Corresponding author.__
*E-mail address: anil.turukmane@vitap.ac.in

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey

As the complexity of the roadway environment in which these vehicles operate continues to increase, it is critical to accurately detect and classify road anomalies to ensure a smooth and safe driving experience. Surface roughnesses such as potholes and cracks are a major problem and lead to an increased risk of accidents and vehicle damage. This reduces the operational effectiveness of self-driving technologies. Also, self-driving cars must comply with strict safety standards. The use of real-time road data while driving could improve the safety of vehicles as it could provide important information for route planning [3].

The number of cars on the streets is increasing and pollution levels are going up, which is leading to the increase in the number of large and small potholes in almost all the cities in the country. In the United States, potholes annually cause an average of 3,597 deaths, an increase of more than 50% in such incidents [4]. These potholes are an indication of poorly maintained roads, and may be symptomatic of more serious underlying structural issues [5]. It takes time and manpower to have a trained structural engineer or inspector perform a thorough inspection. Therefore, different initiatives have been proposed by researchers and transportation industry professionals to develop automated systems for detecting potholes [6]. Potholes are difficult to find and a lot of work. Various detection methods have been developed, such as vision-based systems, 3D reconstruction techniques and vibrational analysis.

Recently, the automated pothole identification technology has achieved a lot using various advanced methodologies [7]. Deep learning algorithms have been especially successful for image-based analysis of road conditions. This progress has allowed for great improvement in road feature detection, classification and segmentation. These advances overcome the limitations of traditional manual inspection methods [8,9]. Deep learning models usually require large amounts of data from both source and target domains to perform well. Smaller datasets can increase the risk of overfitting and decrease the generalizability of the model [10]. As reported, the accuracy rate of a CNN-based YOLO V4 tiny model in the task of pothole identification is 85% compared with human arbitrator, which indicates a strong potential for practical applications [11].

Pothole detection and classification can be done effectively using edge detection methods. It takes pavement images as input data, which allows it to accurately detect and classify the damaged areas [12]. The road condition assessments have been carried out using the YOLO (You Only Look Once) algorithm to detect and classify potholes. Deep learning based methods have shown promising accuracy for pothole identification. For example, Park et al. [13] adopted different YOLO models and conducted an automated detection framework on several image datasets, and obtained high precision for pothole detection. Similarly, Ye et al. [14] improved the standard CNN architecture by adding a pre-pooling layer before the first convolutional layer, which improved the feature extraction ability and the overall detection performance. This direction was followed by Bodhan and Sahu [15] who embedded CNNs in a driverless vehicle prototype which enabled the detection of potholes in real-time while moving. Another good technique for surface anomaly detection is with disparity map techniques. Fan et al. [16] presented an approach based on dense disparity map calibration for accurate classification of damaged and undamaged highway segments. Their later work [17] further refined this by transforming and segmenting road disparity maps for better pothole localization. Fan et al. [18] also proposed an unsupervised segmentation based on the Otsu thresholding technique in a subsequent work, obtaining a clear separation of damaged areas through disparity maps enhancement [19]. The most advanced region-based Convolutional Neural Networks (CNNs) that have been thoroughly tested to evaluate their performance in correctly identifying and locating potholes on road surfaces are state-of-the-art (SoTA) CNNs. Fan et al. proposed a graph attention layer (GAL), which is a modified CNN layer to better represent visual information for semantic segmentation tasks, to further improve the segmentation performance [20]. This layer produces a significant improvement in the segmentation accuracy.

Gao et al. [21] presented a data processing method using an industrial camera as the main implement, integrating pattern recognition and grayscale analysis. In a related study, Yebes et al. [22] focused on the identification of road surface irregularities in real-world scenes by building a large dataset of pothole-labeled images captured by cameras. They also improved object detection using SSD deep neural networks and Faster R-CNN to improve the accuracy of the model. Gupta et al. [23] utilized thermal imaging and proposed a novel method using customized ResNet models, namely ResNet-50-RetinaNet and ResNet-34-SSD. This approach led to high accuracy rates, 91.15% and 74.93% in the two cases.

Lim et al. [5] presented an automatic pothole detection systems based on deep learning, using image datasets from open-source libraries such as Pixabay, Flickr and Google Photos. In their work they trained and tested this dataset with two custom models based on YOLOv2. Chun and Ryu [24] suggested a regression based CNN model for diagnosing road surface defects, using a learning approach to enhance detection accuracy. Yousaf et al. [25] also used computer vision to find and detect potholes. A top-down approach was used to analyze pavement images and recognize risks for drivers. Wu et al. [26] proposed a road pothole detection method based on the fusion of mobile point cloud data and images. This approach uses deep learning algorithm to extract 2D pothole candidates from images, then 3D candidate extraction based on point cloud data. Then a depth analysis is performed for the final pothole detection. Haq et al. suggested an automation approach that integrated block-matching algorithms and key point screening. This technique was a vital step forward for effective roadway pothole detection.

Alhassan et al. [27] proposed a method to enhance the performance of random forest classifiers by incorporating responsive evolution with advanced feature extraction methods. This method was proposed by Lv et al. [28] for random forest classification , and is referred to as adaptive multi-dimensional tree optimization (AMDTO). They also proposed an improved version of YOLOv3, named YOLOv3-ALL, which integrates the K-means++ algorithm with the intersection-over-union (IOU) metric to improve the clustering performance, especially in detecting small features on uneven surfaces. Also, they enhanced network capacity with a convolutional block attention module (CBAM) [29, 30]. Warg et al. [31] improved the ability of the YOLOv3 network to detect faults in small objects by adding a fourth-scale prediction, which improved the detection of small anomalies. In addition, data augmentation and structure optimization of the YOLOv3 model were performed, including color correction and geometric transformations to enhance the data. The potholes were then separated into underwater and dry categories by distinguishing them based on presence or absence of water. The YOLOv3 model was further optimized by combining ResNet 101 and CIoU loss. The data was clustered with K-means++ algorithm and anchor sizes were adapted for multiple scales [32]. To improve the quality of road surface images and to improve the detection of small objects, an advanced method, the enhanced super-resolution generative adversarial network (ESRGAN), was used. The proposed approach used YOLOv5 and EfficientNet detection networks and outperformed the current state-of-the-art methods during the experimental evaluation on several detection datasets.

Deep learning based approaches generally achieve better results than traditional machine learning methods. But there still are limitations and challenges to overcome. Deep learning models process large amounts of data to achieve better results than traditional approaches. Training these complex models can also require substantial financial resources. In addition, deep learning algorithms take more time to train in comparison with other algorithms. Traditional machine learning models are outperformed by transfer learning models because they use features and weights of models trained before, so they own the characteristics of data. This approach accelerates neural network training compared to starting the training process from scratch. In this study, the methodology with the following advancements have been proposed:

- An efficient image preparation technique, incorporating geometric modification and color correction, to improve the pothole, crack, and normal datasets.
- Data from the pre-trained VGG-16 model used as an alternative model for extracting features from pothole images.
- A CNN approach proposed for accurate identification and classification of potholes in road infrastructure.
- Achievement of optimal detection results by enhancing the performance of the model through hyperparameter optimization in the recommended network.
- Comparison of performance of the current model to existing deep learning methods to evaluate its overall effectiveness and quality.

MATERIALS AND METHODS

The algorithm learned from a huge dataset of images. The prediction performance of the proposed CNN architecture was improved by using a pre-trained CNN model VGG-16. This set up allowed for more accurate weights to be created. The parameters and weights of the pre-trained model were trained separately and then transferred to the CNN model to improve its prediction performance. Before combining with other images we applied data augmentation techniques like horizontal and vertical zooming on each image. To get accurate classification results, the ReLU activation function was used in every layer and the SoftMax function was used in the prediction layer. The proposed transfer learning (TL) model used the architecture of VGG-16. A new fully connected CNN layer replaced the last three fully connected (FC) layers of the VGG-16 network to align output characteristics with the requirements of binary classification. Figure 1 shows the data flow within the proposed framework for pothole detection and classification.

Data Collection

Deep learning algorithms can predict future outcomes with reliability and ensure the model's accuracy and robustness. A good training requires a comprehensive dataset of non-conforming images. This implies that there is a need of a sufficient number of datasets with various non-conformities. Rahman and Patel (2020) stated the importance of enough nonconforming image datasets in the findings [33]. We had 6,000 images, which were classified systematically and stored locally. They were divided into three categories: potholes, cracks, and normal roads. The images were 256 pixels by 256 pixels. The images were classified by an expert panel who first identified the presence of potholes, cracks or normal road conditions, and then precisely located these features in each image. To guarantee the full and consistent processing of the dataset, each image was subjected to an object classification algorithm. Each picture showed one or more cracks or potholes. The dataset was composed of the images and their assigned labels. The data set was divided into different subsets for training, testing and validation; 70% for training, 10% for testing and 20% for validation. Figure 2 presents example potholes, cracks, and typical road conditions from the testing subset [34].

Data Preprocessing

The dataset was standardized for its brightness using color correction techniques. Image enhancement techniques were applied in order to increase the clarity and contrast. The quality and quantity of the dataset is a key of the performance of deep learning models. In addition, geometric augmentation techniques were also used to further augment the dataset to ensure robust training and evaluation.

1) Color Correction

The color composition of each pixel typically is determined by three separate color channels: red, green, and blue. These channels were used to detect pavement cracks. The image was represented as a vector composed of all these color channels within a predetermined color space. The color images were processed for contrast enhancement and sharpening. Since the images of the pothole were limited in the grayscale range, the contrast enhancement method was used to increase the grayscale spectrum and improve the clarity of the image. A probability smoothing method was also used to enhance the intensity and saturation of the hue-saturation-intensity (HIS) color model. By standardizing the ranges of the diffusion and concentration components, this adjustment ensured consistent image features. These methods allowed for accurate calculation of the saturation and intensity components. The calculations used are provided in equations (1) and (2) below:

$$y_{1k} = F(x_{1k}) = P\{x_1 \leq x_{1k}\} = \sum_{k=0}^{m} f(x_{1m}) = \sum_{k=0}^{m} P\{x_{1k} \leq x_{1k}\} \tag{1}$$

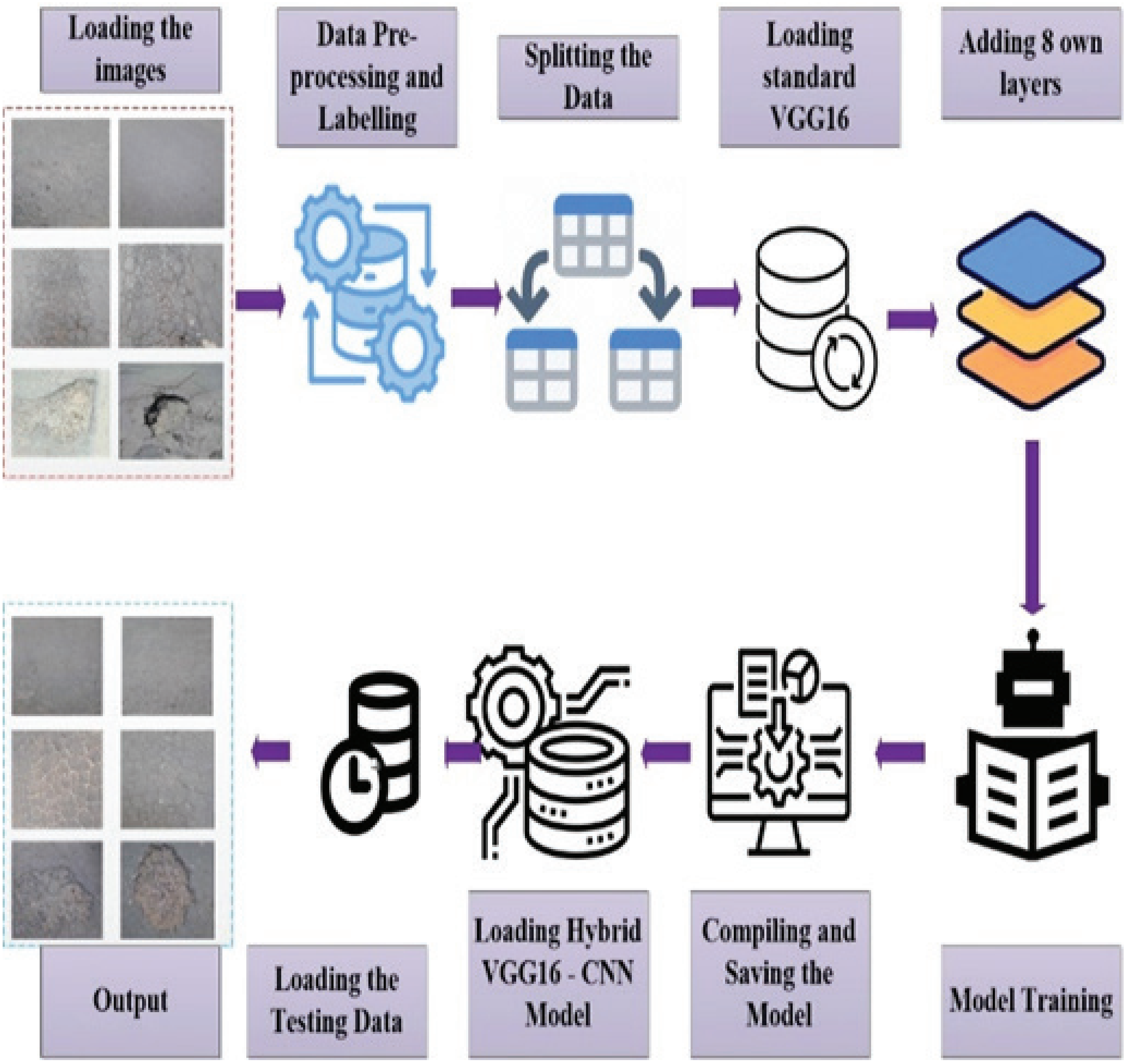

Figure. 1. Process flow diagram of proposed model.

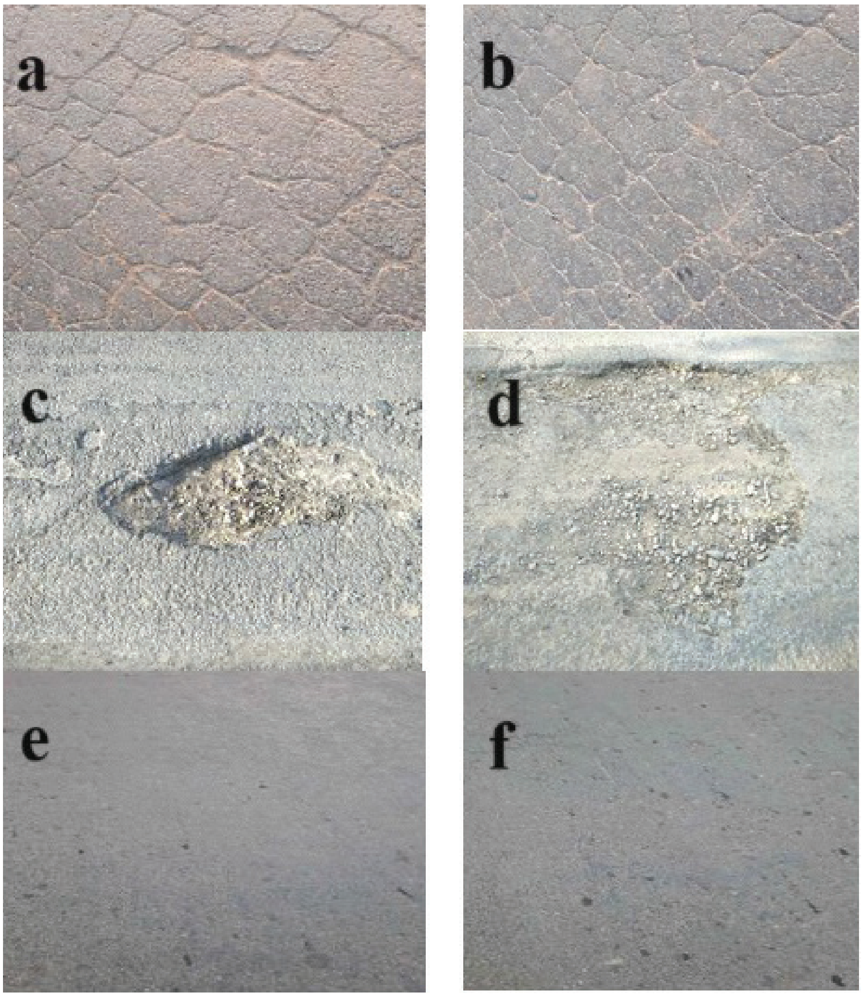

Figure 2. Images depicting cracks (a, b), pothole (c, d) and normal (e, f) road surfaces from the dataset.

$$y_{\mathrm{st}} = F(\, x_{1k} \mid x_{\mathrm{l}k} \,) = \sum_{t=0}^{m} f(\, x_{\mathrm{st}m} \mid x_{\mathrm{l}k} \,) = \sum_{t=0}^{m} \frac{P\{x_{\mathrm{l}}=x_{\mathrm{l}k}, x_{\mathrm{s}}=x_{sm}\}}{P\{x_{\mathrm{l}}=x_{\mathrm{l}k}\}} \quad (2)$$

In this setting k can take values between 0 and L-1 and t can take values between 0 and M-1. Herein, L represents the discrete intensity levels and M represents the saturation, respectively.

$$X = (x_H, x_S, x_I)(xH, xS, xI) \quad (3)$$

T, a vector composed of color pixel values, represents each image. A probability function, denoted by F(.), is defined as follows:

$$F(Z) = F(x_I, x_s) = P(x_I, x_s) \quad (4)$$

By sharpening, the initial blurriness of the pothole image is removed, which increases the contrast of neighboring pixels to highlight the main features of the image. The Laplace operator is used to add the gradient details to pothole image. It improves the sharpness. The enhancement technique by using the Laplace operator is shown in Equation (5).

$$g(x, y) = f(x, y) - [\nabla^2 R(x, y), \nabla^2 G(x, y), \nabla^2 B(x, y)] \quad (5)$$

Here, f(x, y) represents the original pothole image, and g(x, y) represents the sharpened pothole image. Where R(x, y), G(x, y), and B(x, y) are the notations representing the blue, green, and red color channels of the image in the Laplacian operators.

2) Geometric Enhancement

The images were subjected to geometric transformations such as cropping, rotation, and flipping. The transformations of the polygonal regions representing craters were carried out in the uniform and random modes. The four rotation procedures are: 90-degree clockwise rotation, 180-degree counterclockwise rotation, and random crop. Various versions of the potholes images were produced by rotating the original image at a certain angle at the center point. There were various modifications based on different rotation angles. This method effectively augmented the size of the dataset. Random cropping improved the model's ability to better depict the essential features of potholes and increased robustness by cropping edges of the image randomly. In particular, potholes (P) are features (F) in the presence of background noise (N). The main feature of the captured pothole image is represented as (F, N). While the model can theoretically learn (F, N), it is more inclined to learn F, which can cause overfitting. The majority of the pothole images are in the middle position, which decreases the chance of missing important features as opposed to background noise (N). This positional arrangement significantly adds to the information gain of feature F, favoring its learning over N, as indicated in equation (6). An example of this is the expression Xc(1) (1.F,0.B), which assigns a weight of 1 to F and a weight of 0 to B.

$$X_I^{(1)}(1.F,0.N), X_I^{(2)}(1.F,0.N), X_I^{(3)}(1.F,0.N) \quad (6)$$

3) Relativistic Discriminator

The purpose of the relativistic discriminator (D_R), in contract to the ordinary discriminator (D), is to calculate the likelihood that a real image will be noticeably more realistic than a generated image. This method allows for precise calculation of adversarial loss and discriminator loss in the convolutional layer.

Adversarial training was used to improve the performance of the generator in super-resolution (SR) tasks, where the generator was trained on both real and synthetic data to generate realistic images. This process included averaging the outputs over a mini-batch to stabilize learning and calculate expected values like the adversarial loss. The generator minimized the difference between generated samples (x, f) and real data (x, r), which greatly improved the ability to generate high-quality SR images. Then the generator improved its output by using gradients from real and generated data and improved its functionality and output quality over the time. The calculations were done using the following equations:

$$L_{D_R} = -E_{x_r}\left[\log\left(D_R(x_r, x_f)\right)\right] - E_{x_f}\left[\log\left(D_R(x_f, x_r)\right)\right] \quad (7)$$

$$L_{G_R} = -E_{x_r}\left[\log\left(1 - D_R(x_r, x_f)\right)\right] - E_{x_f}\left[\log\left(D_R(x_f, x_r)\right)\right] \quad (8)$$

Here, Exf represents the process of averaging over all data produced within a mini-batch. The outputs were derived from the input low-resolution (LR) images.

4) Perceptual Loss

SRGAN suggested a loss function to restrict feature representations after activation. Similar to perceptual loss Lpercept, feature constraints were applied before activation. The encoder network loss can be calculated as follows:

$$L_G = L_{percept} + \lambda L_R^G + \eta L_1 \quad (9)$$

The L_1 norm is used to measure the 1-norm distance between the original images and the reconstructed images. This method ensures that the loss function effectively preserves the integrity of the content. In addition, a network interpolation method is proposed, which combines a pre-trained network optimized for peak signal-to-noise ratio (PSNR) with a network fine-tuned using generative adversarial networks (GANs). This technique aims to construct an integrated model, Gg, that contains optimal parameters by merging the parameters of both networks. The following equations describe the process of combining different loss functions to create an integrated model that balances multiple objectives in the training of a generative model.

$$\Theta_{INT}^{ERP}, \Theta_{PSNR}^{G}, \Theta_{GAN} \quad (10)$$

$$\Theta_{INT}^{ERP} = (1-\alpha)\Theta_{PSNR}^{G} + \alpha\Theta_{GAN} \quad (11)$$

The network was able to generate high quality results without introducing any artifacts and was able to maintain a stable perceptual quality during the training process. It also effectively reduced image noise, performed precise deblurring and enhanced the edge details of potholes. The images were also zoomed to four times their original size.

Model of transfer learning

Transfer learning : The process of taking a pre-trained model and fine-tuning it with new datasets to improve performance on specific tasks. Figure 3 shows how this method exploits the learned features and capabilities of the model. The pre-trained model is a feature extractor that learns essential information such as edges and shapes. It is widely used in image classification, object detection and other tasks. In transfer learning, the model layers are pre-trained for feature extraction, if the new task is similar to the original dataset and the complexity of ImageNet. Transfer learning helps to lessen the training of parameters and lessen the size of the dataset needed to obtain a good CNN model. With transfer learning, only the fully connected layers are replaced and trained with the data of the specific task. This is especially useful for fields like radiology, where data sets can be sparse but model performance is critical.

The pothole detection model was trained on the transfer learning framework based on VGG-16 as shown in Figure 3. It took the features learned by the pre-trained VGG-16 model. The pre-trained model was employed with its bottom layers as static feature extractors, gathering universal traits and not altering them during training. The top layers were changed to extract task-specific information from the pothole image dataset. The VGG-16 model was modified by replacing the fully connected layers with CNN layers and processing the output with a SoftMax layer. For the binary task, the output layer was a single node that outputted either a 1 (pothole) or 0 (no pothole).

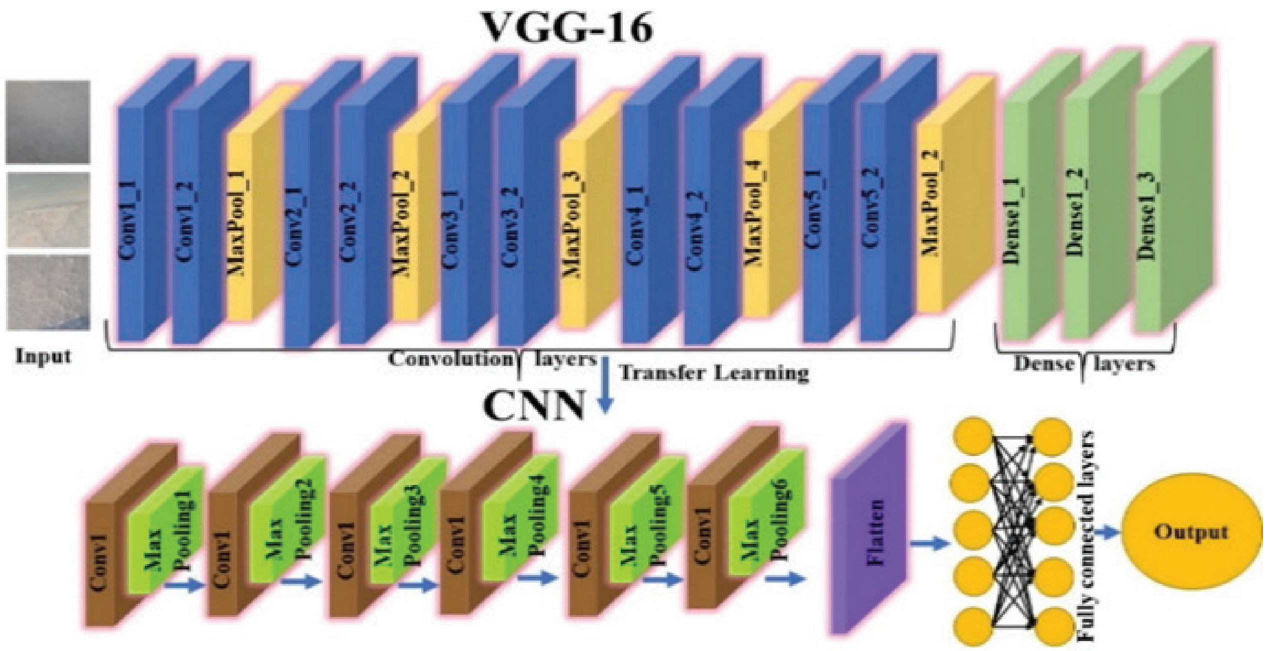

Figure 3. Architectural design of the proposed model.

Hyperparameter Optimization

Tuning hyper-parameters is one of the most important steps in building a reliable and broadly-applicable model. The configuration of hidden layer units in the CNN, batch size, number of epochs, learning rate, activation function selection, dropout rates for regularization, and other critical parameters must be carefully tuned to enhance the model's ability to generalize to new data. The best hyperparameters were selected within the set of possible values, considering the impact of each parameter on the model's performance. The selection of the optimizer, the batch size and the number of training epochs were important factors during the hyperparameter tuning. In modern deep learning, Adam optimizer has been widely used instead of the Stochastic Gradient Descent (SGD) optimizer, which combines the benefits of RMSProp and AdaGrad algorithms. The mathematical formulation of the Adam optimizer is given in Eqs. (12) and (13).

$$p_t = \alpha_1 p_{t-1} + (1-\alpha_1)\left[\frac{\delta L}{\delta W_t}\right] \tag{12}$$

$$q_t = \alpha_2 q_{t-1} + (1-\alpha_2)\left[\frac{\delta L}{\delta W_t}\right] \tag{13}$$

Herein, optimization parameters, α1 and α2 represent decay rates, δL denotes the derivative of the loss function and δW_t signifies the derivative of the weights at iteration t. WHere, t denotes the current weights, p_t denotes the gradients and q_t denotes the sum of squared past gradients.

Batch size is an important hyperparameter in deep learning systems. GPUs have the capability of parallel processing, which allows for larger batch sizes to speed up computation during model training. However, this acceleration can lead to reduced generalization performance if not managed correctly. Smaller batch sizes may help to achieve faster convergence to optimal solutions but might require more iterations to train effectively. Thus there is a trade off between the costs and benefits of using larger versus smaller batch sizes. The model was trained with batch sizes of 8, 16, 32, 64, and 128 for this study. The whole data set was given to the neural network for each " epoch " . The internal parameters of the model were trained using the training sample and then completed the epoch. The number of epochs has been increased to reduce the errors made during the learning process, but more epochs need more computational time. Therefore, it is important to find the best compromise between more or less epochs. In this study, a total of 55 epochs were used. All relevant information for this investigation is provided in Table 1, which is the exhaustive list of hyperparameters and their corresponding values.

Table 1. Hyperparameters and their values

Hyperparameters	Values
Batch size	64
Learning rate	0.0001
Optimizer	Adam
Activation function	ReLU
Epoch	55

RESULTS AND DISCUSSION

Simulation experiments were performed on the Anaconda Navigator 2.5.2 platform with Python 3.9. The model was trained and tested on Windows machines. The CNN model was built using TensorFlow version 1.13 and Keras version 2.2.4 was used for the core API and for the high-level API. The model converged at epoch 55. The images were divided into 70% for training, 20% for validation and 10% for testing. The proposed model was validated with several assessment measures to evaluate its feasibility and performance. We tested it on a notebook with Jupyter, Python 3.9, a GPU, and 8GB of RAM. The CNN model was constructed using TensorFlow version 1.12 as the backend and Keras version 2.2.4 as the high-level API. To ensure consistency of results, the test was repeated several times and the data from each repetition was gathered and graphed. The proposed model was tested on the dataset and the performance measures such as F1 score, accuracy, precision and specificity were studied. The hyperparameters were exhaustively searched and different deep learning models like MobileNetV2, Xception, VGG16 were experimented to find out their ability to detect unexpected road condition.

Performance Metrics

During the tests, several criteria were used to evaluate the performance of the models as discussed below

Accuracy

When evaluating the success of deep learning algorithms, it is common to focus on a single metric, known as a measure. This is often due to its simplicity and ease of implementation. Accuracy is typically described as the ratio of correctly classified cases to the total number of cases and is calculated by the following equation:

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \tag{14}$$

When a case is correctly categorized as positive, it is referred to as a true positive (TP), and when it is correctly classified as negative, it is called a true negative (TN). Conversely, a false negative (FN) refers to cases that are incorrectly labeled as negative, while a false positive (FP) refers to cases that are incorrectly labeled as positive.

Specificity

In classification models, specificity measures the percentage of true negative cases that the classifier correctly identifies. It is given by the following equation:

$$\text{Specificity} = \frac{\text{TN}}{\text{TP+FN}} \tag{15}$$

Precision and recall

Due to their interrelationship, precision and recall are often considered together. The following formulas can be used to represent precision and recall mathematically:

$$Precision = \frac{TP}{TP+FP} \tag{16}$$

$$Recall = \frac{TP}{\text{TP+FN}} \tag{17}$$

F1 Score

F1 Score is a quantitative metric that combines precision and recall into a single metric by calculating their harmonic mean. It balances both measures and gives you one value you can use. It is the best metric to measure the overall performance of a classification system , especially when you have imbalanced data . F1 score is useful in cases where the cost of false positives and false negatives is high. It combines precision (number of correct positive predictions) and recall (number of positives found) into a single metric. The F1 score is calculated using the following equation.

$$F1\ score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \tag{18}$$

F1 score is often used since it provides a good overview about precision and recall. This helps to understand if adjustments need to be made or if the algorithm has to be re-checked. This statistic is useful because it gives a fair measure of recall and accuracy. Nevertheless, a weighted average taking into consideration both criteria necessitate care while implementing it.

Performance Assessment of the Proposed Model

Accuracy and loss measures were used in both the training and validation stages of the proposed VGG-16-CNN model to assess its performance. Lower loss values indicate higher model performance. However, the interpretation of loss metrics depends on the performance of the model on the training and validation datasets. Figure 4a exhibits the model training and validation accuracy during 50 epochs. At the start of the first epochs, both the training and validation accuracy increase quickly, suggesting that the model is quickly learning the data patterns and improving predictions. With training, the accuracies of both datasets converge with less and more gradual improvements after about 10 epochs . You can see that both curves converge at around 96 % accuracy by about epoch 30 . This is a good sign that the model has learned to perform well . The closeness of the training and validation curves indicates that there is little over-fitting, as the model performs similarly on both data sets. This pattern shows the model is able to generalize well to unseen data and maintain strong accuracy across various epochs.

Figure 4b shows the training and validation loss of the model for 50 epochs. Both curves are steadily decreasing, which shows consistent improvement in the model performance. In the first 10 epochs, training and validation loss decrease significantly, showing that the model learns the data quickly and reduces error efficiently. After this first phase , the loss reduction slows down , and the two curves

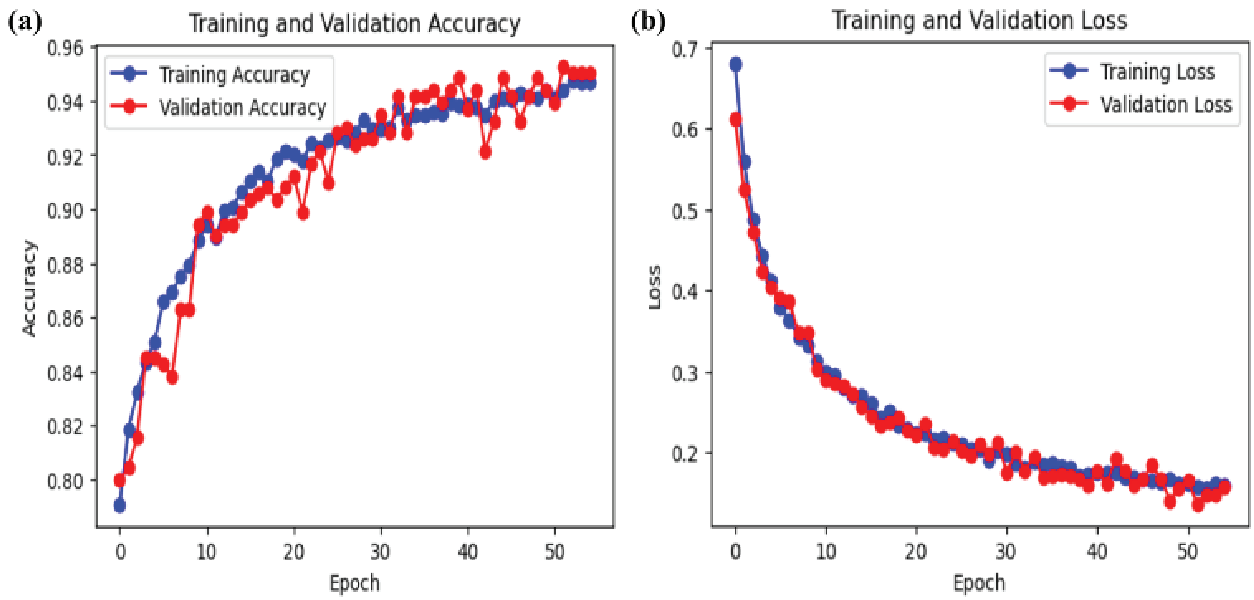

Figure 4. **(a)** Accuracy of the model during training and validation phases, **(b)** Loss of the model during training and validation phases.

start to stabilize . Both training and validation loss values tend to stabilize at low values by around epoch 30, indicating that the model has reached a state of optimized learning. The proximity of training and validation loss curves shows very little over-fitting and hence the model is generalized to unseen data and robust in performance.

Figure 5a presents the specificity comparison between the predicted classes using the proposed model. The high specificity of the model demonstrates its capacity to categorize most negative events accurately. More data and better image quality increase the generalization of the model, by increasing data variability and reducing the risk of overfitting. The proposed method achieved specificity of 98.6% for the Normal category, 95.5% for the Crack category and 96.9% for the Pothole category. Figure 5b shows the precision achieved by the proposed model. The results obtained are for Normal 97.3% , Crack 91.1% and Pothole 93.6% which proves the superiority of the proposed method over the existing algorithms. Image enhancement techniques were used to improve the pothole images and the photographs became clearer and more visually interpretable. This enabled accurate pothole detection and reduced cognitive load for image interpretation. Figure 5c illustrates the recall scores. The recall performance of the proposed algorithm is compared with the state-of-the-art methods. The proposed method achieves 95.2% recall for Normal, 90.3% recall for Crack and 93.0% recall for Pothole. Image enhancement improves the accuracy of pothole detection by emphasizing important features, reducing noise, and enhancing image details. Experimental results show that the learning effectiveness of the algorithm can be improved by the improvement of data quality and image clarity through image enhancement methods. The hyperparameter optimization is also required as it optimizes the deep learning algorithm so that the prediction and classification of potholes are accurate.

Figure 5d shows the F1 scores of the proposed method and the existing methods. The proposed transfer learning model in this study achieved significantly better results with F1 scores of 96.2% for Normal, 90.7% for Crack, and 93.3% for Pothole. Data augmentation is a very important step in balancing the imbalanced datasets and improving the performance and obtaining best results which was not possible only with the original dataset. Figure 5e shows the average macro scores of all performance parameters. The overall accuracy is 93%, with macro precision at 94%, macro recall at 92.8%, macro specificity at 97%, and macro F1 score at 93.4% for the proposed approach.

Two statistical tests, the Matthews Correlation Coefficient (MCC) and Cohen's Kappa, were used to evaluate and compare the results. MCC assesses binary classification performance by focusing on accurate predictions of both positive and negative cases. It serves as a key metric for evaluating binary classification accuracy and is represented by equation (19) in the research literature as follows:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{19}$$

MCC is generally in the range of -1 to +1 . Values close to +1 indicate excellent performance . The proposed model achieved MCC of 0.8934 which indicates a very strong correlation between the predicted images and the actual images

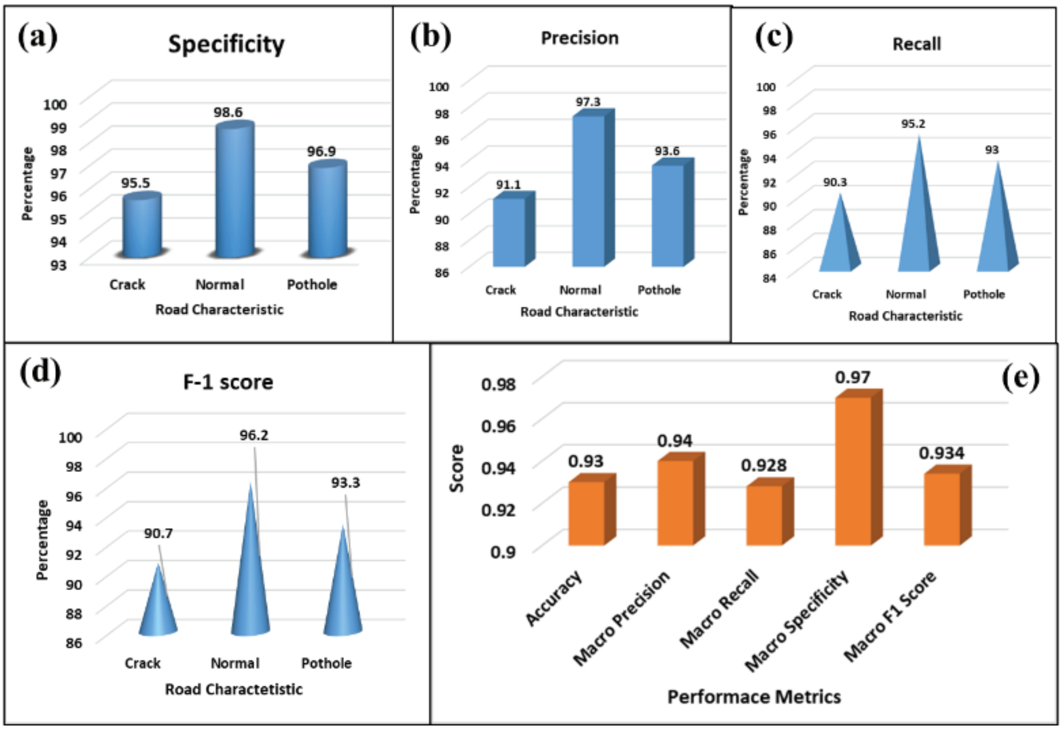

Figure 5. Performance metrics of the proposed model comparing its performance in cracked normal and pothole road characteristics: (a) Specificity, (b) Precision, (c) Recall, (d) Macro F1 score, (e) Comparison of all metrics.

in the dataset. This high value of MCC confirms that our model predicts much better than the random guessing. It also means balanced precision and recall, that is, a model that does a good job minimizing both false negatives and false positives.

Cohen's Kappa (CK) Statistic measures the level of agreement between two raters who categorize items into mutually exclusive groups. This agreement is represented mathematically in equation (20).

$$CK = \frac{p_0 - p_e}{1 - p_e} \tag{20}$$

This explains the agreement level in the raters' observations. (pe) is the probability of agreement expected by chance between two randomly chosen items. The formulas (21) and (22) are used to calculate (p_0) and (p_e) in the raters. The formulas (23) and (24) are used.

$$p_0 = \frac{TP+TN}{TP+TN+FP+FN} \tag{21}$$

$$p_e = \textit{Positive of Probability of Positive} + \textit{Positive of Negativity} \tag{22}$$

$$\text{Probability of Positive} = \left(\frac{\mathrm{TP+FP}}{\mathrm{TP+TN+FP+FN}}\right) \times \left(\frac{\mathrm{TP+FN}}{\mathrm{TP+TN+FP+FN}}\right) \tag{23}$$

$$\text{Probability of Negative} = \left(\frac{FP+TN}{TP+TN+FP+FN}\right) \times \left(\frac{FN+TN}{TP+TN+FP+FN}\right) \tag{24}$$

CK is usually between 0 and 1. A score of 1 corresponds to complete agreement between two assessors and a score of 0 corresponds to complete disagreement. The coefficient CK can be at most equal to 1 and it represents the degree of agreement between the real data and the prediction models. All CK models are in almost perfect agreement, indicating a high degree of consistency in the projections. The CK value of 0.896 for the proposed model shows that 89.6% of agreement is achieved in detecting anomalies on the road.

The performance evaluation of preprocessed data includes the evaluation of data warping and repair rates to measure the performance of the model. The difference of the pixel value between raw image and processed output is the mean squared error (MSE). This makes it possible to evaluate the distortion of the original data before preprocessing and the degree of its successful correction. It is calculated by the following equation:

$$\mathrm{MSE} = \frac{1}{AB}\sum_{m=0}^{A-1}\sum_{n=0}^{B-1}[I(m,n) - N(m,n)]^2 \tag{25}$$

The MSE value 0.0315 of the proposed model proves that there is a very small deviation between the input and predicted output images. The pre-trained CNN-VGG-16 model is used for feature extraction and transfer learning on the dataset. The model is used to classify images of cracks, potholes and normal road surfaces. The model is then trained on a specific dataset, using the large amount of data available, to improve its predictive accuracy. Then its performance is evaluated on a separate testing data set and hence predictions can be improved based on the knowledge gained during training. In an A×B image, data loss due to image compression can alter the original data. The Peak Signal to Noise Ratio (PSNR) is a metric used to quantify the difference between the original information and the noisy or distorted signal (Equation 26). This metric measures the information loss in the image and is the ratio of interference to the maximum possible signal intensity. The PSNR value is decreased means the image quality is reduced with minimal loss of signal fidelity. PSNR cannot be calculated for lossless images where the mean square error is zero.

$$PSNR = 10\log_{10}\left(\frac{\text{Max}^2}{\text{MSE}}\right) \quad (26)$$

The proposed model shows good accuracy in distinguishing between pothole and non-pothole classes from the assessment results with error levels for the non-pothole class varying from 0.02 to 0.03 (MSE). A higher PSNR (lower MSE) is expected in higher quality datasets than in lower quality datasets, because higher quality datasets are related to lower MSE (see Max2). As demonstrated by PSNR values greater than 50 dB, the proposed method efficiently minimizes the image loss and provides accurate pothole predictions. The PSNR of our model is 63.147dB and the MSE is 0.0315 for Max 255, which shows the efficacy of the model for potholes detection in different environmental conditions [35].

ROC Curve

The ROC curve illustrates the variation in sensitivity and specificity at different testing thresholds. This test measures the classifier's ability to discriminate between positive and negative examples. The closer the shape is to the top left corner, the better the performance of the algorithm. ROC curve is an important tool for issue detection and classification. It compares the true positive rate (TPR) to the false positive rate (FPR) to measure the ability of the model to distinguish between real signals and noise [36]. This gives a comprehensive evaluation of how the model performs. Sensitivity, or true positive rate (TPR), measures the ability of the model to detect positive cases. Specificity, on the other hand, measures the ability of the model to not classify instances incorrectly as positive. The Area Under the Curve (AUC) is a scale-dependent measure of the overall correctness of the test, varying from 0 (completely wrong) to 1 (perfectly correct). The ROC measures the model's ability to distinguish between different classes, with higher values indicating better performance. The area of the ROC curve (Fig. 6) closer to the top left corner highlights the successful identification of potholes by the proposed method. The classifier accuracy in finding positive and negative cases, i.e. distinguishing between cracks, potholes and normal road conditions, was evaluated using the ROC curve. Figure 7 shows the performance of the classifier. The AUC values-0.89 for class 0 (crack), 0.95 for class 1 (normal) and 0.92 for class 2 (pothole) indicate the high accuracy of the classification of cracks, potholes

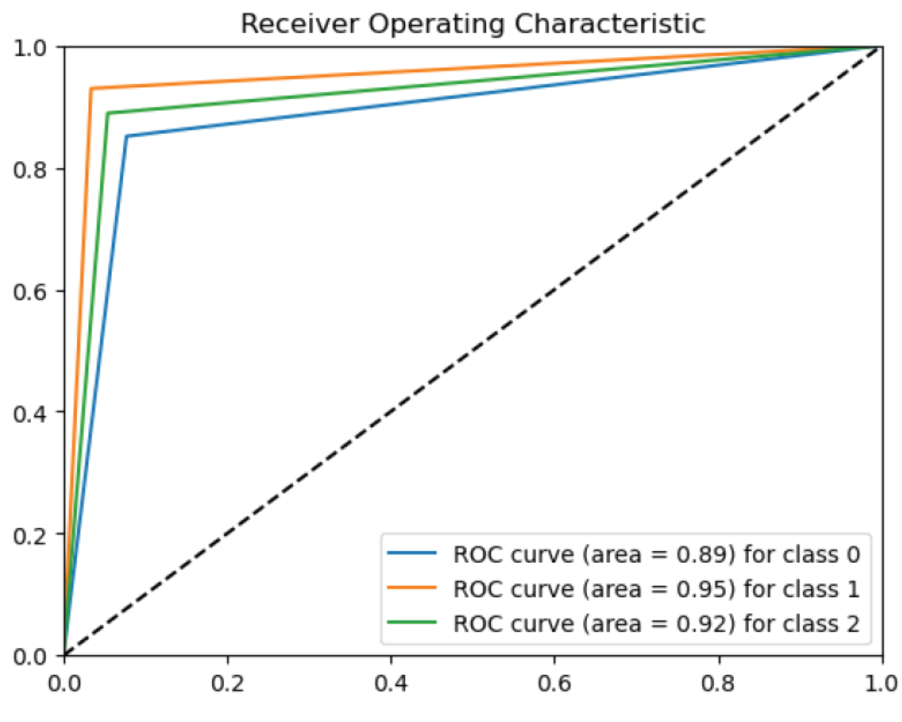

Figure 6. ROC graph of my proposed model

Table 2. Comparison of our planned work to others published work

Authors	Accuracy (%)	Precision (%)	Recall (%)
Azher et.al. [37]	87	86.5	87
Koch et.al. [38]	85.9	81.6	86.1
Ryu et.al.[39]	73.5	80.0	72.3
Shi et.al.[40]	92.2	90.0	90.0
Dihao et.al.[41]	92.0	91.7	90.0
Gopal Krishnan et. al.[42]	90	90	90
Current work	**93**	**94**	**92.8**

and normal road conditions. These results show that the model is very effective in detecting cracks and potholes.

Comparison of the proposed model with existing literature

Table 2 presents a comparison of accuracy, precision, and recall for the detection of road anomalies between our proposed work and previously published studies [37]. The results show that our proposed approach achieves superior performance in accuracy (93%), precision (94%), and recall (92.8%). However, it is important to note that the models used in the comparison study were trained under different settings. For example, all the models for road anomaly detection were trained on datasets with varying image quality, dataset size, and GPU specifications. These differences can significantly impact the results.

A closer examination of the findings reveals some inconsistencies that could affect the interpretation of model performance. Factors such as inference speed and model size are fundamental aspects of object recognition models, which remain relatively constant regardless of the efforts of developers. It is worth noting that the road anomaly detection model in our work is better than many other lightweight models, though it is not as powerful as more complex and resource-consuming models. This means that there is a balance between model complexity and performance, and our approach has a good balance for road anomaly detection.

CONCLUSION

In this work, a robust deep learning model for road anomaly detection is proposed. Their model is based on a VGG-16 based CNN with transfer learning. The model performed very well in terms of several metrics, including accuracy, specificity, precision, recall, and F1 score. The model achieved a high overall accuracy of 93% with an excellent precision and recall, surpassing some existing methods for pothole and crack detection. The training and validation accuracy and loss curves indicate good generalization and no overfitting of the model, which was also confirmed by the consistent results in multiple runs.

The performance metrics analysis showed specificity scores of 98.6% for Normal, 95.5% for Crack and 96.9% for Pothole, and a strong F1 score which showed the balanced approach of the model to precision and recall. The MCC and Cohen's Kappa coefficients further confirmed the accuracy and consistency of the model's predictions with values of 0.8934 and 0.896, respectively. The high values confirm the model's capability to minimize false positives and false negatives, ensuring reliable detection of road anomalies. The application of image enhancement techniques was very useful for the model to detect potholes by improving the clarity of input images, which in turn improved prediction accuracy. More specifically, transfer learning was utilized, exploiting the pre-trained VGG-16 model, to efficiently extract features, subsequently fine-tuned on the target dataset, thereby enhancing the robustness and predictive power of the model.

Compared to other state-of-the-art methods, the proposed approach not only achieved superior performance in terms of different performance metrics but also exhibited the advantage of a lower computational cost, making it more feasible for real-time applications. The proposed model is a promising solution for road anomaly detection. Future work could include improving the performance of the model by incorporating additional data sources and fine-tuning the hyperparameters for improved accuracy. Although our model performs well for the day-time conditions, future work can explore low-light adaptation through infrared sensors or data augmentation to make it robust in the night-time or in bad weather conditions such as heavy rain. These adaptations will improve its performance and make it suitable for real-time use in autonomous vehicles operating in different environments.

ACKNOWLEDGEMENTS

The authors sincerely thank the Coordinator of the High-Performance Computing Laboratory at VIT-AP University, Amaravati, Andhra Pradesh, India, for his generous support and for granting access to advanced computational facilities. His assistance played a vital role in enabling the experimental analysis conducted in this study.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Alqethami S, Alghamdi S, Tsubait T, Alhakami H. Roadnet: Efficient model to detect and classify road damages. Appl Sci 2022;12:11529. [CrossRef]

[2] Wu H, Yao L, Xu Z, Li Y, Ao X, Chen Q, et al. Road pothole extraction and safety evaluation by integration of point cloud and images derived from mobile mapping sensors. Adv Eng Inform 2019;42:100936. [CrossRef]

[3] Varona B, Monteserin A, Teyseyre A. A deep learning approach to automatic road surface monitoring and pothole detection. Pers Ubiquitous Comput 2020;24:519−534. [CrossRef]

[4] Bhatia Y, Rai R, Gupta V, Aggarwal N, Akula A, et al. Convolutional neural networks-based potholes detection using thermal imaging. J King Saud Univ Comput Inf Sci 2022;34:578−588. [CrossRef]

[5] Kim YM, Kim YG, Son SY, Lim SY, Choi BY, Choi DH. Review of recent automated pothole-detection methods. Appl Sci 2022;12:5320. [CrossRef]

[6] Chen H, Yao M, Gu Q. Pothole detection using location-aware convolutional neural networks. Int J Mach Learn Cybern 2020;11:899−911. [CrossRef]

[7] Dib J, Sirlantzis K, Howells G. A review on negative road anomaly detection methods. IEEE Access 2020;8:57298−57316. [CrossRef]

[8] Sun Z, Pei L, Li W, Hao X, Chen Y. Pavement encapsulation crack detection method based on improved faster r-cnn. J South China Univ Technol (Nat Sci Ed) 2020;48:84−93.

[9] Liyang X, Wei L, Zhaoyun S, Pei L. Automatic crack detection method based on JTG 5210-2018 standard. In: Proc, Maintenance and Management, Branch of China Highway Society, China. 2020. p. 90−98.

[10] Ellingson SR, Davis B, Allen J. Machine learning and ligand binding predictions: A review of data, methods, and obstacles. Biochim Biophys Acta G Gen Subj 2020;1864:129545. [CrossRef]

[11] Hossain MS, Angan RB, Hasan MM. Pothole detection and estimation of repair cost in Bangladeshi street: AI-based multiple case analysis. In: International Conference on Electrical, Computer and Communication Engineering (ECCE), IEEE. 2023. p. 1−6. [CrossRef]

[12] Baek JW, Chung K. Pothole classification model using edge detection in road image. Appl Sci 2020;10:6662. [CrossRef]

[13] Park SS, Tran VT, Lee DE. Application of various YOLO models for computer vision-based real-time pothole detection. Appl Sci 2021;11:11229. [CrossRef]

[14] Ye Wei, Jiang W, Tong Z, Yuan D, Xiao J. Convolutional neural network for pothole detection in asphalt pavement. Road Mater Pavement Des 2021;22:42−58. [CrossRef]

[15] Dewangan DK, Sahu SP. Potnet: Pothole detection for autonomous vehicle system using convolutional neural network. Electron Lett 2021;57:53−56. [CrossRef]

[16] Fan R, Liu M. Road damage detection based on unsupervised disparity map segmentation. IEEE Trans Intell Transp Syst 2019;21:4906−4911. [CrossRef]

[17] Fan R, Wang H, Wang Y, Liu M, Pitas I. Graph attention layer evolves semantic segmentation for road pothole detection: A benchmark and algorithms. IEEE Trans Image Process 2021;30:8144−8154. [CrossRef]

[18] Fan R, Ozgunalp U, Hosking B, Liu M, Pitas I. Pothole detection based on disparity transformation and road surface modeling. IEEE Trans Image Process 2019;29:897−908. [CrossRef]

[19] Devendiran R, Turukmane AV. Dugat-LSTM: Deep learning based network intrusion detection system using chaotic optimization strategy. Expert Syst Appl 2024;245:123027. [CrossRef]

[20] Tanwar N, Turukmane AV. Comparative analysis of multiple methods utilized for road pothole identification using deep learning. Doctoral Symposium on Computational Intelligence. Singapore: Springer Nature Singapore. 2024;1086:531−540. [CrossRef]

[21] Gao M, Wang X, Zhu S, Guan P. Detection and segmentation of cement concrete pavement pothole based on image processing technology. Math Probl Eng 2020;2020:1−13. [CrossRef]

[22] Yebes JJ, Montero D, Arriola I. Learning to automatically catch potholes in worldwide road scene images. IEEE Intell Transp Syst Mag 2020;13:192−205. [CrossRef]

[23] Gupta S, Sharma P, Sharma D, Gupta V, Sambyal N. Detection and localization of potholes in thermal images using deep neural networks. Multimed Tools Appl 2020;79:26265–26284. [CrossRef]

[24] Chun C, Ryu SK. Road surface damage detection using fully convolutional neural networks and semi-supervised learning. Sensors 2019;19:5501. [CrossRef]

[25] Tanwar N, Turukmane AV. Modified MobileNetV2 transfer learning model to detect road potholes. PeerJ Comput Sci 2025;11:e2519. [CrossRef]

[26] Haq MUU, Ashfaque M, Mathavan S, Kamal K, Ahmed A. Stereo-based 3d reconstruction of potholes by a hybrid, dense matching scheme. IEEE Sens J 2019;19:3807–3817. [CrossRef]

[27] Liu J, Luo H, Liu H. Deep learning-based data analytics for safety in construction. Autom Constr 2022;140:104302. [CrossRef]

[28] Lv N, Xiao J, Qiao Y. Object detection algorithm for surface defects based on a novel YOLOv3 model. Processes 2022;10:701. [CrossRef]

[29] Swain S, Tripathy AK. A novel smart data collection approach in UAV-enabled smart transportation system. In: Advances in Distributed Computing and Machine Learning: Proceedings of ICADCML 2022. Springer; 2022. p. 709-716. [CrossRef]

[30] Babbar S, Bedi J. Real-time traffic, accident, and potholes detection by deep learning techniques: A modern approach for traffic management. Neural Comput Appl 2023;35:19465-19479. [CrossRef]

[31] Wang D, Liu Z, Gu X, Wu W, Chen Y, Wang L. Automatic detection of pothole distress in asphalt pavement using improved convolutional neural networks. Remote Sens 2022;14:3892. [CrossRef]

[32] Tanwar N, Turukmane AV. Effective hyper-parameter tuning of machine learning model for analysis drinking water quality. Grenze Int J Eng Technol 2024;10:10-17.

[33] Liu Z, Gu X, Wu W, Zou X, Dong Q, Wang L. GPR-based detection of internal cracks in asphalt pavement: A combination method of deepaugment data and object detection. Measurement. 2022;197:111281. [CrossRef]

[34] Turukmane AV, Devendiran R. M-MultiSVM: An efficient feature selection assisted network intrusion detection system using machine learning. Comput Secur 2024;137:103587. [CrossRef]

[35] Devendiran R, Turukmane AV. Dugat-LSTM: Deep learning based network intrusion detection system using chaotic optimization strategy. Expert Syst Appl 245;2024:123027. [CrossRef]

[36] [Dib J, Sirlantzis K, Howells G. A review on negative road anomaly detection methods. IEEE Access 2020;8:57298-57316. [CrossRef]

[37] Azhar K, Murtaza F, Habib HA. Computer vision based detection and localization of potholes in asphalt pavement images. In: 2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE). IEEE; 2016. p. 1-5. [CrossRef]

[38] Koch C, Brilakis I. Pothole detection in asphalt pavement images. Adv Eng Inform 2011;25:507-515. [CrossRef]

[39] Ryu SK, Kim T, Kim YR. Image-based pothole detection system for ITS service and road management system. Math Probl Eng 2015;2015:968361. [CrossRef]

[40] Shi Y, Cui L, Qi Z, Meng F, Chen Z. Automatic road crack detection using random structured forests. IEEE Trans Intell Transp Syst 2016;17:3434-3445. [CrossRef]

[41] You-quan H, Jian W, Jian-fang X. A research of pavement potholes detection based on three-dimensional projection transformation. In: 2011 4th International Congress on Image and Signal Processing. IEEE; 2011. p. 1805-1808. [CrossRef]

[42] Gopalakrishnan K, Khaitan SK, Choudhary A, Agrawal A. Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection. Constr Build Mater 2017;157:322-330. [CrossRef]