

Research Article

Identification of Civil Infrastructure Damage Using Ensemble Transfer Learning Model

A. Shamila Ebenezer ¹, S. Deepa Kanmani ², V. Sheela ³, K. Ramalakshmi ⁴,
V. Chandran ⁵, M. G. Sumithra ⁵, B. Elakkiya ⁶ and Bharani Murugesan ⁷

¹Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India

²Department of Information Technology, Sri Krishna College of Engineering and Technology, Coimbatore, Tamil Nadu, India

³Department of Civil Engineering, Sri Ranganathar Institute of Engineering and Technology, Coimbatore, Tamil Nadu, India

⁴Department of Computer Science and Engineering, Alliance School of Engineering and Design, Alliance University, Bangalore, Karnataka, India

⁵Department of Electronics and Communication Engineering, Dr. N. G. P. Institute of Technology, Kalapatti, Coimbatore 641048, India

⁶Department of Electronics and Communication Engineering, Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Chennai, Tamil Nadu 600062, India

⁷School of Textile Leather and Fashion Technology Kombolcha 208, Kombolcha Institute of Technology, Wollo University, South Wollo, Ethiopia

Correspondence should be addressed to Bharani Murugesan; bharani.murugesan@kiot.edu.et

Received 20 February 2021; Revised 25 March 2021; Accepted 28 September 2021; Published 20 October 2021

Academic Editor: B. B. Gupta

Copyright © 2021 A. Shamila Ebenezer et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article uses cutting-edge deep learning technology to identify structural damage from images for a civil engineering application. The public infrastructures of the country are generally inspected physically by a visual evaluation by qualified inspectors. However, manual inspections are pretty time-consuming and often require too much labor. The number of experts capable of evaluating such structural damage is inadequate. As a result, computer vision-based techniques for automatic damage detection have been developed. This paper's civil infrastructure damages are classified into four damages of roads common in Indian highways and the concrete deterioration in the bridges. The convolutional neural network has become a standard tool for organizing and recognizing images. In this paper, an ensemble of three CNN models is proposed, and two are transfer learning-based models. The proposed ensemble transfer learning model provided a validation accuracy of 87.1%.

1. Introduction

Bridges, bridges, highways, lifeline networks, and houses are only a few examples of critical infrastructure that was constructed decades ago and is no longer fit for purpose. As per the American Society of Civil Engineers' 2017 Infrastructure Report Card, the USA has over 56,000 operationally defective bridges that would cost an additional 123 billion dollars to patch [1]. Data collected by evaluation and assessment processes is often used to determine the state of

structural design. Conventional methods for monitoring the effects of structural design comprise routine examination by qualified inspectors along with specific decision-making criteria. Such inspections, on the other hand, can be time-consuming, labor-intensive, costly, and hazardous [2]. Monitoring can be used to get a quantitative understanding of a structure's current condition by measuring the physical quantities like excitations, pressures, and lateral displacement; such methods allow for seamless concrete structures surveillance in real time, with the expectation of achieving

safety and performance, as well as decreased quality control costs [3]. Although these approaches have been shown to yield accurate findings, they usually have low visual acuity or require massive sensing devices to be installed. Another problem is that when the sensors are integrated, access to them is often limited, causing device management to be difficult. Contact sensor installation is challenging and time-consuming when periodic monitoring is required [4]. To advance and understand the benefits of automated civil infrastructure condition assessment, improved inspection and monitoring approaches with less human interference, lower cost, and higher spatial resolution must be developed and tested. Government agencies expend substantial amounts of money conducting routine inspections of essential infrastructure such as bridges and roads [1].

Furthermore, systemic response records and photographs as data media are becoming increasingly important in today's data explosion, especially in deep learning (DL) applications in computer vision, which have made significant progress in recent years. Furthermore, the application of machine learning and deep learning aims to make computers perform labor-intensive repetitive tasks while also learning from previous experiences [2]. In recent days, vision-based structural health monitoring is used to identify structural damage detection, which heavily relies on the experience of the human to do a visual inspection. In this proposed work, the civil infrastructure damages considered are concrete deterioration in the bridges, road cracks, road potholes, road rutting, and road lines [3]. First, the concrete deterioration in the bridges is considered due to a reduction in the lifespan of the reinforcement steel. Bridge collapses are equivalent to plane crashes and natural disasters resulting in injury, loss of life, and property damage [4]. Common bridge damages are corrosion, leaching, cracks, and scaling of the concrete surface to identify conditions; the bridge structures are exposed to a series of destructive tests including cutting, sewing, and taking core samples. It helps in analyzing the core profile and testing the bridges. However, coring increases the risk of damaging an already damaged structure due to the removal of concrete by cutting, sewing, and core sample. Destructive testing methods are not ideal for locating the bridge structures' damage comprising a larger area [5].

Secondly, the road network in India extremely covers a larger area to connect various places through roadways. However, the maintenance of the roads is so poor. As roads contribute indirectly to the country's economic development, the roads must be well laid out and solid. India being a developing country, the need for well-maintained civil infrastructure is growing. Being the largest country with the highest population, this problem is yet to be addressed in India. Therefore, numerous solutions for automated road damage inspection, including 3D methods, vibration-based techniques, and vision-based techniques, were provided by researchers [6]. The vibration detection techniques are limited to the contact road section. While thermal imaging approaches provide detailed information on road damages, there is a need to

close the door. The 3D hologram images are generated using quaternion Fresnel transform of civil infrastructure to analyze its condition [7].

In the meantime, image processing techniques are cost-effective. However, it can be hampered by a lack of accuracy. Road damages are one of the root causes of road accidents and need to be maintained appropriately to avoid dangerous accidents. A pothole is caused due to the damage in the asphalt sheet internally, which affects the surface of the pavement [8]. Owing to the poor soil quality, extreme temperature fluctuations create such potholes. Hair, edge joint, lane joint, widening, edge, perception, evaporation, alligator, and erosion are the nine distinct types of cracks found in the road's superficial layer. The cracks in the roads may lead to delay in travel time and increased costs due to excessive fuel consumption and vehicle maintenance [9]. Early detection of such cracks and other damages is essential to avoid the severe problems caused due to pavement failures. Deep convolutional neural networks (CNNs) are useful in a variety of applications such as computer vision [10], natural language processing [11], image processing [12], speech recognition, signal classification, and battery estimation [13]. This study's key contribution can be summarized as follows:

- (1) The new dataset of the total of 1176 damages consists of 133 concrete deterioration cases, 483 roadways potholes, 360 roadway cracks, 100 roadway rutting cases, and 100 roadway lines from a different source open-sourced in GitHub
- (2) This is the first time that a deep neural network model is developed to categorize both bridge deterioration and road damages to the best of our knowledge
- (3) To extract high-level features from the data and shape the feature map to identify the damages, the custom shallow CNN models are explicitly designed
- (4) The Xception and AlexNet model is used as a transfer learning model due to the less available data for training
- (5) The ensemble model is developed using the custom CNN, Xception, and AlexNet to improve the prediction rate and achieve better performance using the different sets of feature maps
- (6) To verify the algorithm's accuracy and validity, the proposed models are compared with various methods

The remaining part of the paper is structured as follows. Section 2 is a literature review in which we look at other current techniques that prompted us to pursue this path. Section 3 presents materials and methods which elaborate the dataset, data augmentation, and region-based segmentation. Section 4 offers the proposed work to discuss custom CNN, transfer learning, and an ensemble model. Section 5 illustrates the result and discussion of the different state-of-the-art models with the proposed model followed by the conclusion in Section 6, which summarizes this paper.

2. Related Works

Maintenance of the infrastructure should be carried out regularly and extensively. Moreover, it is expensive in labor and material requirements. Therefore, infrastructure monitoring systems allow people to monitor and exchange information on problems with local infrastructures, such as graffiti, broken pavement slabs, and potholes and focus on nonworking street lights by using a similar set of images. Local municipal governments presently have the accountability of monitoring and maintaining all of their facilities. Many local governments in countries cannot carry out adequate examinations due to a lack of economic and human resources. To address the issues concerning infrastructure maintenance, these challenges must be solved as the local government controls most infrastructure. In a real-world situation, when road surface supervisors from a government play a major role to fix the harm of roads, they need to understand, with clarity, what kind of harm is involved and they need to take appropriate action. This section provides a detailed review of the existing techniques for road damage detection. Civil infrastructure and its damage detection are as follows. This experiment focused on the histogram and extracted the image region function and then added a nonlinear SVM kernel to detect the target. The results revealed that the pothole could be easily seen in this analysis.

In [14], the author proposed a neural network-based method for sensing harm using modal properties is introduced, which can take into account the pretending faults in the limited component-based model from which training designs are to be produced. The feasibility of the proposed approach is given two numerical comparisons on a simple beam and a multigirder bridge. Using a backpropagation-based neural network, the researcher demonstrated a method for determining the damage intensities in parts in truss bridge systems [15]. The technique used to solve problems related to several unidentified limitations in a substantial structural design is identifying the substructures. The natural dynamic characteristics and vibration frequency were provided to the neural network as input constraints for classifying the damage, primarily in irregular modal analysis measurements. The machine learning approach's recent growth has given researchers more input to project different algorithms to classify and categorize the road harms using images of the road surface. In [16], the author introduced a method for restructuring the road summary using vehicle responses. An artificial neural network-based method was developed to rebuild the road's surface area from the speed of vehicles. In this method, numerical procedures were created using different rates based on the irregularity grades, increasing defects for varying noise levels, and loads of cars. But this approach is particular for specific applications. The author proposed laser scanning and imaging of the pavement [17]. Automatic detection of cracks was implemented using a backpropagation-based neural network (BPNN) [18]. An image processing framework was developed to preprocess the images and mine the information about cracks. This framework was used to eliminate redundant

information during the initial stage of image processing and applied feature parameters to categorize the gaps in the road images and differentiate linear and alligator or crocodile cracks. Japan road association-maintained road surface safety monitoring was primarily founded on graphical evaluation by professional road surface supervisors and objective determination by broad-scale inspection. This framework can address different large-scale images.

In [19], the author made a study analysis, rather than using a smartphone accelerometer. Researchers aim to achieve automated assessment of road surface conditions by taking images of the road surface using a mobile phone camera. To resolve the identification of the injury, several methods for measuring road surface damage from road surface images have been suggested. The author suggested using a commercial black box camera to locate potholes [20]. The pothole detection structure collects the data, such as location, appearance, and height using the camera. Damage occurs on the road surface, classified into nine crack forms. Cracks in the lane joint, hairline cracks, cracks in the edge, reflective cracks, widening cracks, shrinkage concrete cracks, joint cracks in the borders, and crocodile cracks are also called alligator cracks, crescent, or slippage cracks. These cracks may occur due to unstable soil roads. Their sewage system is not in good condition; there are poor execution of the job and lousy paving materials, which cannot tolerate the road's capacity and traffic [21]. However, labor-intensive approaches, including thorough inspection, require a substantial amount of time and effort. Additionally, the labor-intensive method appears to be confused and unmanageable; hence, it implies a more significant risk. Though detailed inspection, such as MMS and scanning via laser, is highly efficient, thorough inspection requires too much expense. A detailed survey is done, and an application is built using a lightweight road manager smartphone [22].

An Android-based approach finally incorporates the model parameter at every investigation and classifies the road infrastructure images. Cracks and their different types need different kinds of maintenance and repair mechanisms images were gathered, and numerous procedures were applied to extract the significant features to highlight the cracks. The logarithm-based transformation was used to make the darker cells expand and brighter cells compress. Initially, preprocessing was done on images. Later, the decision tree-based heuristic algorithm was executed on the preprocessed images to classify and detect the crack on the photos. We compared the accuracy and runtime speed of state-of-the-art object detection methods using convolutional neural networks to train the harm detection model with our dataset using a GPU server and a smartphone. Finally, we show that the type of damage can be accurately classified into eight categories using the proposed object detection method (33A vibration-based system using the mobile accelerometer) and gyroscope is developed for automatic pavement distress detection, as opposed to a vision-based approach using video processing, which is proposed in this work. The vibration-based device has an accuracy of 80% in detecting potholes, patches, and bumps. For detecting gaps, potholes, and patches, the vision-based approach has

an accuracy of 84% [23]. This study used smartphone sensors and onboard diagnostic equipment to detect roadway pavement irregularities, which resulted in a lower roadway infrastructure assessment. In the research, smartphone sensors and artificial neural network techniques were used to capture a vehicle's contact with a roadway pavement as it travels and then use the observed interaction patterns to locate potholes in the pavement. The device has a detection accuracy of around 90% [24]. A new method for automatically detecting and segmenting pavement cracks from 2D photographs is described in this paper. The proposed method begins by creating a crack attribute vector with a steerable matched filter, which improves the contrast between cracks and surrounding pavement while also capturing crack discontinuity and curvature. The crack saliency map is used to build a coarse crack area and rough crack property estimates. After that, the fine crack area is fed into a region-based active contour model, which is then used to segment cracks using a level set evolution method [20]. This paper describes the process for automatically detecting and classifying road damage, such as potholes and cracks. Damages such as potholes, cracks, and persistent depressions are framed in this method [25]. DeepCrack is a model that is used to detect cracks automatically using the high-level representation is proposed in this paper. To capture line structures, a convolution filter is fused to extract the multi-level features. DeepCrack achieves an average of 0.87 f -score on the three daunting datasets [26]. The overview of the literature survey with advantages and disadvantages is provided in Table 1.

3. Materials and Methods

3.1. Dataset Collection. To build and validate the proposed method, the dataset was created from scratch. To highlight the generality of the proposed solution, photographs from various sources relevant to civil infrastructure were helpful. The damage databases currently available do not include Indian roads and bridge structures. The authors photographed most of the damaged buildings, and photographs of the damages in designs available on the Internet in the public domain were also used. Photos obtained from Google image searches are among the other sources used. The images collected from diverse sources were hand-labeled by the authors. The total number of images is 1176, containing 133 concrete deterioration, 483 potholes images, 360 crack images, 100 rutting images, and 100-line images. The dataset's information is described in Table 2. Figure 1 shows some of the dataset's sample photos.

3.2. Data Augmentation. Deep learning neural network provides better accuracy when trained with a vast dataset. To improve the network's performance, the training data are augmented using the different image augmentation techniques from the existing data using the Keras deep learning. The newly created images belong to the same class as that of the original image. Image augmentation using transforms such as horizontal flipping, vertical flipping, zooms, rotating,

clipping, and shifts was performed to enhance and expand the dataset with new images. Although deep learning algorithms can learn features invariant to their location in the photo, the data augmentation can aid the model in learning components that are also invariant to transforms, such as light levels in photographs, ordering, and more. Some of the image augmentation results are shown in Figure 2, where the horizontal flipping, vertical flipping, clipping, and 90° rotation are performed.

3.3. Region-Based Image Segmentation. Segmentation of the additional damages was done based on the pixel values. The cracks were segmented based on multiple threshold values. The pixel-based segmentation is less complex as the computations are more straightforward and produce faster results. Pixel-based image segmentation is shown in Figure 3 which shows the image after pixel-based segmentation.

4. Proposed Deep Learning Framework

The proposed deep learning framework is shown in Figure 4. The data obtained are preprocessed. In the preprocessing stage, the image segmentation technique is used to extract useful features from the input image. The preprocessed images are fed into the three models where the first model is developed from scratch. The second model employed the pretrained (AlexNet and Xception model). The third model is the ensemble model of all three models to improve the proposed work's performance. The ensemble model uses the majority voting scheme to predict the output.

4.1. Convolutional Neural Network. Convolutional neural networks (CNNs) are a form of feed-forward neural network used primarily for image classification. A CNN is made up of 3 layers: a convolutional layer, a pooling layer, and a fully connected layer. In general, images are fed into convolutional layers in a CNN. After the operations in the "convolutional kernel," the output is fed into the pooling layers for downsampling, where the convolutional production is a feature chart. The primary part of CNN is a layer used to extract the features from the given input images. The process of forwarding propagation in each kernel k is represented in

$$a_k^l = \sigma(z_k^l) = \sigma(a^{l-1} * W_k^l + b^l), \quad (1)$$

where $\sigma(z_k^l)$ is the activation function, W_k^l represents the convolutional weight kernel, b^l represents the bias in layer l , and a_k^l represents the output of kernel k from layer l . The output convolutional layer could be computed as follows:

$$a^l = \sigma(z^l) = \sigma\left(\sum_{k=1}^M z_k^l\right) = \sigma\left(\sum_{k=1}^M (a_k^l * W_k^l) + b^l\right). \quad (2)$$

Followed by the convolutional layer, the pooling layer is used to reduce the feature dimensionality. The pooling function may be max, min, or average. However, max-pooling is widely used as it reports maximum output in the output field. The activation function is the most important

TABLE 1: Overview of the literature survey.

S. No	Ref	Model used	Merits	Demerits
1	[1]	CNN Model	This model detects the damage location and damage severity with a validation accuracy of 89.06%	(i) But the model is implemented with the static condition (ii) The bridge is temporarily closed and determined using static loading conditions (iii) This can be expanded for the dynamic loading condition
2	[2]	Fast R-CNN	Data augmentation gives a better mean <i>F1</i> -Score of 6.255 when compared with R-CNN 6.174	Much weightage is given for the preprocessing techniques model evaluation training is not discussed
3	[14]	Neural network-based damage detection method	Efficient method for pattern detection	(i) A considerable amount of false alarms (ii) Structural damages of the bridges are not identified properly
4	[19]	FCN-based road surface anomaly detection model	The road anomalies and the location detection can be given approximately	But the depth of the anomalies cannot be determined by the proposed model
5	[27]	A black box camera is used to detect a pothole	(i) Potholes detection (ii) Suitable for pavement management system	(i) Black box camera uses a limited computing environment. The intensity of the light affects the false detection
6	[28]	Decision Tree heuristic algorithm	Pavement crack detection	(i) More processing time due to preprocessing of images (ii) Horizontal and vertical classification of pictures needed for further crack detection
7	[29]	SSD inception V2 and SSD MobileNet	Developed for classifying eight types of road damages	(i) Could achieve a precision of 77% and recall of 71% only (ii) Dataset consists of comparatively fewer pothole images
8	[30]	Vibration sensor-based method	It can be implemented using smartphones with sensors	(i) The results are with low accuracy (ii) The vehicle's suspension and speed affect the amplitude vibration and the accuracy
9	[24]	Machine learning (ML)-based approach	Less computational and time complexity compared to deep learning models	(i) Hand-picked feature extraction using traditional image processing methods (ii) Machine learning models produce less accurate results than deep learning models
10	[20]	Image processing-based crack detection	Cost-efficient and fast	(i) There is only one form of damage found (ii) This leads to a high error rate in the presence of poor illumination and distortion
11	[25]	Machine learning (ML)-based approach	Less computational and time complexity compared to deep learning models	(i) Only for pothole detection (ii) The results of machine learning models are comparatively lower than the deep learning models
12	[26]	Encoder decoder architecture of SegNet DeepCrack	This can be used for edge detection and images and vessel detection in retina images	The accuracy can be improved; it produces a 0.87 <i>F</i> -measure

TABLE 2: Dataset distribution.

Damage type	Number of images
Concrete deterioration in bridges	133
Roadways potholes	483
Roadways cracks	360
Roadways rutting	100
Roadways lines	100
Total	1176

part of CNN because it is responsible for deciding the network's output and computational performance. It also has a considerable effect on convergence speed. There are different activation functions such as sigmoid, tanh, ReLU, LReLU, PReLU, Swish, and softmax. Rectified Linear Unit is

the most widely used activation function in deep learning applications. It provides the advantages of training an algorithm considerably fast by converging the weight updates quickly. Generally, in the output layer, the sigmoid function is used for binary classification, and softmax is used in

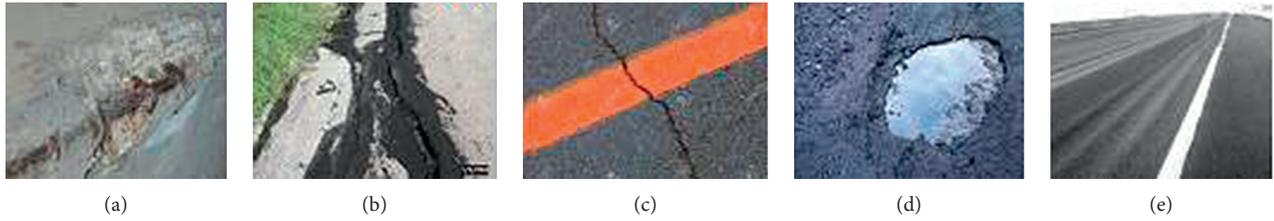


FIGURE 1: Sample images from the dataset: (a) bridge damage, (b) roadways cracks, (c) roadways lines, (d) roadways potholes, and (e) roadways rutting.

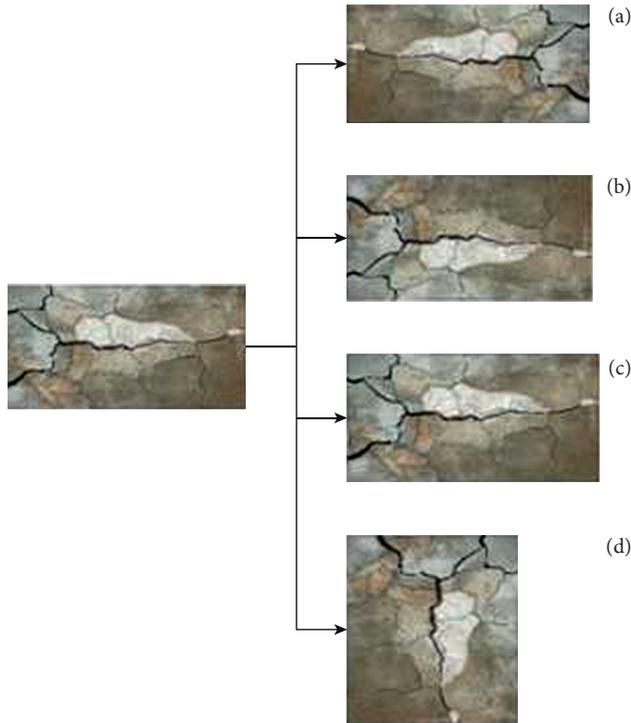


FIGURE 2: Augmented images: (a) horizontal flipping, (b) vertical flipping, (c) clipping, and (d) rotating.

multiple types. The softmax function is used in the output layer of the proposed CNN model, which is defined by the following equation for n numbers $x_1 \dots x_n$ is given in

$$s(x_i) = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}}. \quad (3)$$

The output ranges from 0 to 1 and adds up to 1 at the prediction stage.

4.2. Custom CNN Model. To have robustness in the model, the convolutional neural network model was developed from scratch by considering that the intense network may lead to longer training time, and a minimal network may result in poor learning. By considering these issues, the custom CNN model is proposed, and the structure is mentioned in Table 2. The input image size is resized to 64×64 width and height, respectively. The convolutional layer has a fixed kernel size of 3×3 , but with a different number of filters with 16, 32, and 64 as the depth of the

architecture progresses. The number of filters increased to extract the different sets of features. The flattening layer is used on the fully connected layer to convert a 2-dimensional image to a single-column vector. The softmax layer is used as the classification layer to classify the five different sets of classes. The optimization function used in the model is Adaptive Momentum Estimation (ADAM) to avoid local minima [31]. ADAM optimization function also supports adaptive learning using the network parameters. The categorical cross-entropy loss function used in the model as the problem is a multiclass classification problem. Table 3 summarizes the structure of the proposed network. The overall architecture is shown in Figure 5 to classify the damages in the civil infrastructure.

4.3. Transfer Learning Model. Transfer learning is a technique where the pretrained models are used instead of developing a scratch model. Transfer learning is where the pretrained models are trained on the ImageNet dataset consisting of millions of images. It can be used in feature extraction from the dataset followed by the flattened layer of CNN, and also, the pretrained models can be used as the classification network. In this proposed work, the Xception model is used for feature extraction. The Google inception architecture inspires the Xception model. The inception paradigm is interpreted in an “extreme” way in Xception. The Xception architecture is a depth-wise separable convolution layer stack with residual connections that are linearly stacked. Xception architecture uses depth-wise convolution instead of regular convolutional. The depth-wise convolutional reduced the training time by significantly reducing the parameter count. The Xception model is used as the feature extraction and used as the transfer learning model and given as input to the fully connected layers. It is shown in Figure 6. The output layer uses the softmax activation function to classify the given five classes.

The transfer learning model is also used in the AlexNet model, where it is used as the function extractor unit. Alex Krizhevsky had suggested it. It is a basic CNN structure made up of convolutional and ultimately linked layers. The input dimensions are 227×227 pixels. The first convolution layer receives the data. There are four more convolution layers after that. In the convolution layers, the activation feature used is ReLU. The softmax activation feature is used in the output layer. The AlexNet architecture for image classification is shown in Figure 7.

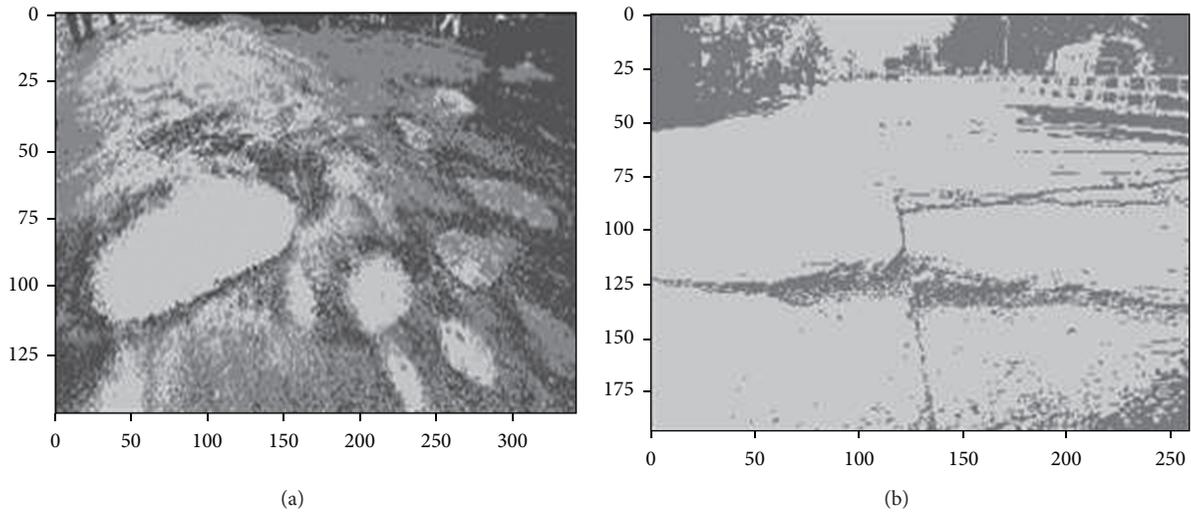


FIGURE 3: Pixel-based image segmentation of infrastructure images: (a) damaged roads and (b) cracked roads.

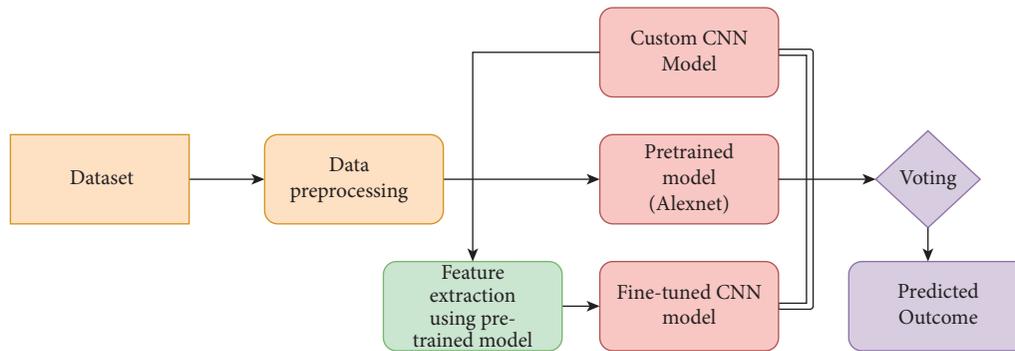


FIGURE 4: Overall framework of the proposed work.

5. Results and Discussion

The experiment was performed in the Google Colab GPU runtime environment platform using Keras and Tensorflow. The three convolutional network architectures' performance is compared with the ensemble CNN architectures and other state-of-the-art architectures for classifying the civil infrastructure. The entire dataset is split into the training and testing ratios of 90% and 10%. The training data contains 1027 images, and the validation data includes 153 images. Later, the training data is processed for data augmentation to enhance the training dataset. The accuracy of the model with feature extraction using pretrained models like Xception and AlexNet is analyzed. The four models are trained and tested with the obtained dataset for multiclass classification. The hyperparameter fixed for the proposed models is a learning rate of 0.001, an epoch of 100, and batch size of 8, and an ADAM optimizer is used. The training data is used to train the custom CNN model, which is then tested with validation data. The data augmentation and pixel-based segmentation features are used to avoid overfitting in the model. Despite efforts taken to avoid overfitting by the regularization techniques, it was observed that the proposed custom CNN

reached a convergence state between the scale of 80% and 86% of classification accuracy and began to memorize training data. The highest classification accuracy was observed at the 36th epoch as 86% for the validation dataset and it is shown in Figure 8. The model weight and bias from the 36th epoch are taken for model testing. Xception model is trained on the obtained dataset to classify bridge damages and road damage classification and obtained the training accuracy 96% and validation accuracy of 82% shown in Figure 9. The gap between the training and validation accuracy is quite high due to the dataset class imbalance problem. Similarly, for the AlexNet model, the training accuracy ranges between 92% and 96%, and validation accuracy is 84% due to the same class imbalance problem and interclass similarity in the dataset shown in Figure 10. In order to solve this problem, the ensemble model is developed from the three models by combining the architecture of custom CNN, AlexNet, and Xception models. The pre-processed dataset is fed into all three architectures at the same time in order to extract discriminative features from the feature map, and then, the majority voting scheme is used to distinguish the various types of damages. For the obtained dataset shown in Figure 11, the ensemble approach

TABLE 3: Structure of custom CNN model.

Layer	Layer type	No. of nodes	Kernel size	Activation	Dropout rate
0	Input shape	64×64	—	—	—
1	Conv1	16	3×3	ReLU	—
2	Conv2	16	3×3	ReLU	—
3	Max-pool	—	2×2	—	0.2
4	Conv3	32	3×3	ReLU	—
5	Conv4	32	3×3	ReLU	—
6	Max-pool	—	2×2	—	0.2
7	Conv5	64	3×3	ReLU	—
8	Conv6	64	3×3	ReLU	—
9	Max-pool	—	2×2	—	—
10	FC	512	—	ReLU	—
11	Output	5	—	Softmax	—

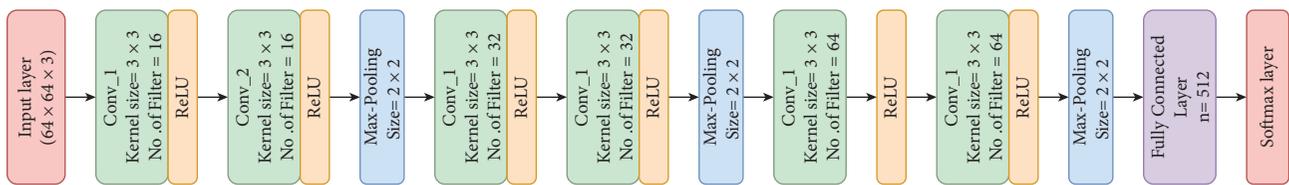


FIGURE 5: Custom convolutional neural network for classifying civil infrastructure damages.

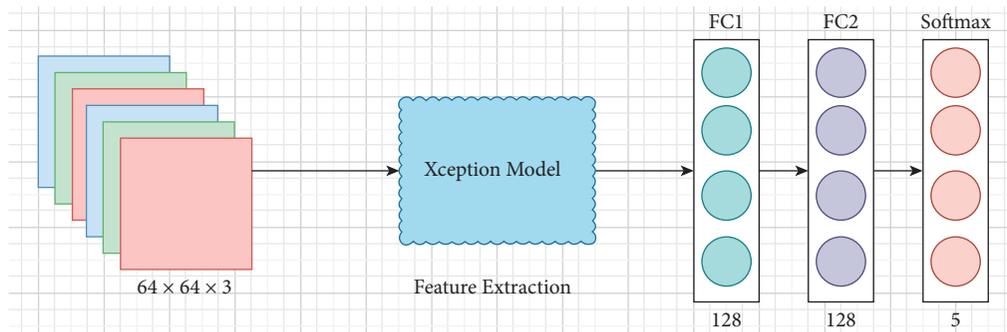


FIGURE 6: Xception model as transfer learning model used for feature extraction.

provides a training accuracy of 94% and a validation accuracy of 87%.

To assess the model's robustness, all of the proposed models are compared to various state-of-the-art architecture. The proposed method is the only method trained and tested for various classes such as roadways cracks, lines, potholes, and rutting and bridge deterioration detection. But the model proposed in the literature is trained and tested with specific damages, either road damages for crack or pothole and bridge damage detection; they were not specific to Indian roads. The performance analysis of the different models with the proposed method is shown in Table 4. Accuracy is the only metric discussed in most of the techniques discussed in the literature. The precision and recall values calculated from the confusion matrix are performed with a deeper understanding of the model. The actual positive value, false-positive value, true negative value, and false-negative value are calculated. The accuracy is the metric that defines how well the model performs in predicting the authentic positive classes and actual hostile

classes over all the types in the dataset. The precision metric is defined as the ratio of true positives to all positive predicted values. The recall metric is defined as the ratio of true positives to all actual positives.

Table 4 in the Resnet model is presented to classify the pothole in the road and achieves an accuracy of 90.5% of classifying pothole and nonpothole using thermal photos [32]. From the confusion matrix discussed in the literature, the precision and recall are calculated as 87.2% and 80.2%. The traditional image processing techniques have been used in classifying the pothole in the road, as proposed in [33]. The overall accuracy of the model is achieved as 73.5%. The precision and recall of the method are 80.0% and 73.3%, respectively. The proposed ensemble model provides an accuracy of 87% for the obtained dataset. It also achieves better precision and recall of 84.92% and 83.53% compared to other proposed models in this work. The Resnet [32] model and ANN [34] in the table provide better accuracy when compared to all other models because the models were trained on the dataset to classify potholes or nonpotholes in

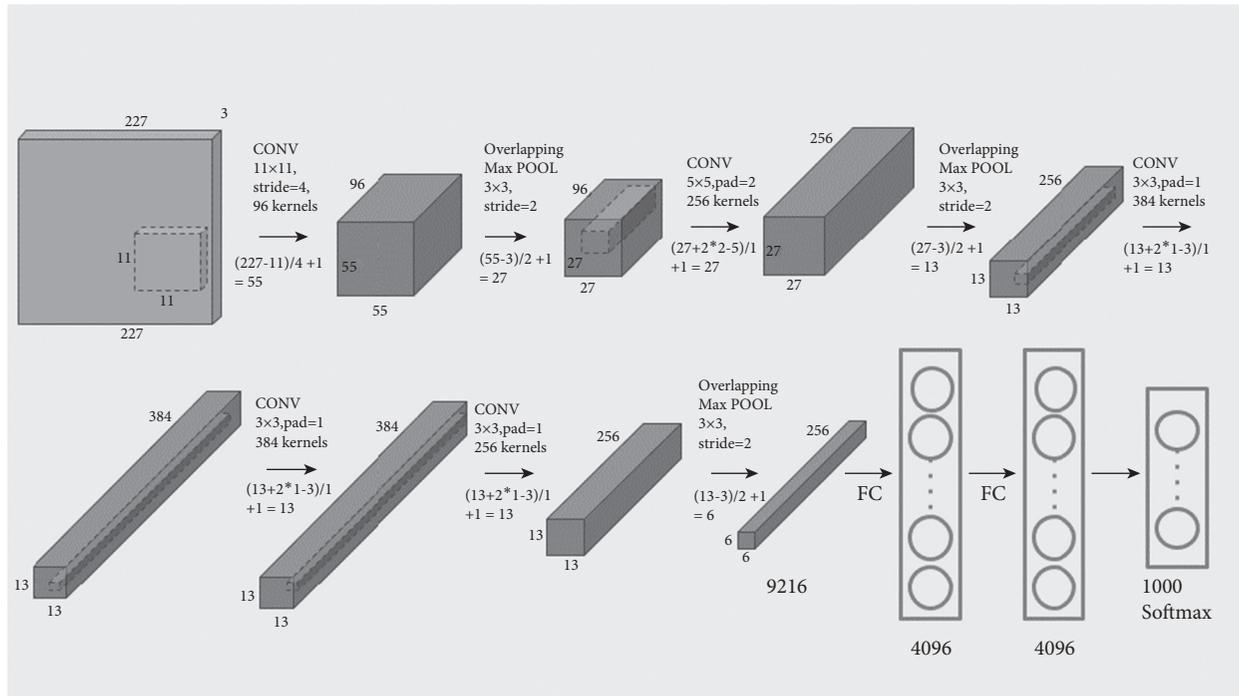


FIGURE 7: AlexNet model as transfer learning model used for feature extraction.

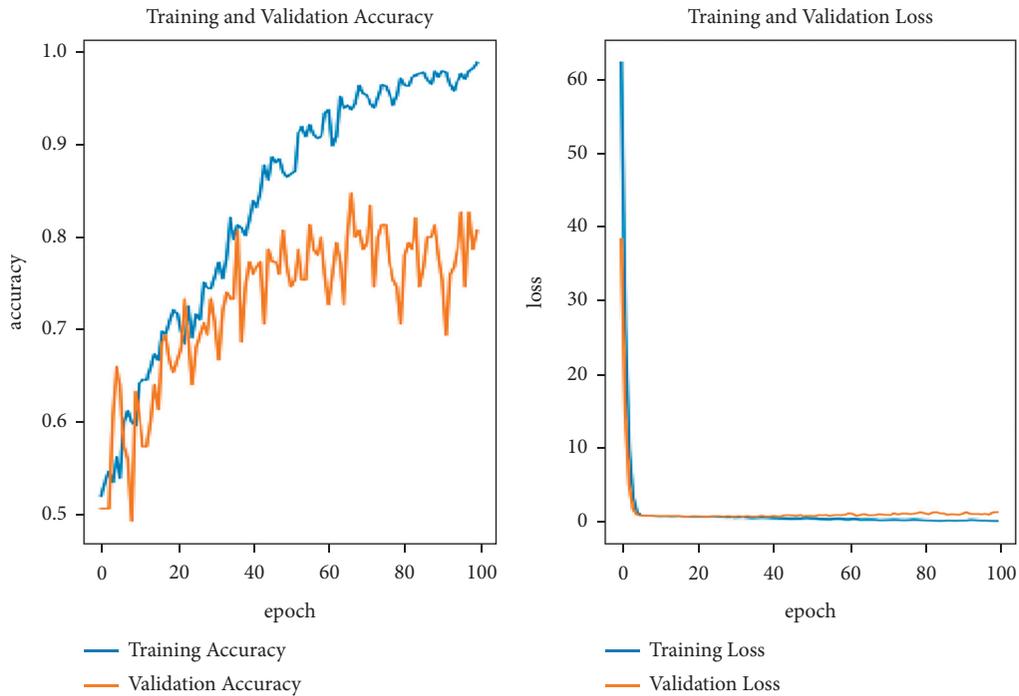


FIGURE 8: Accuracy and loss of the custom CNN model.

the road. Road damage is detected using CNN-based methods and achieves an accuracy of 81.4% [23]. But the proposed method is trained on the multiclass dataset with different inter- and intraclass similarities. The graphical representation of the above-discussed process is shown in Figure 12.

The overall runtime of the other pretrained models is provided in Table 5. The training time is calculated in seconds for each epoch of the model trained on the obtained dataset. By considering the best three models, the AlexNet, Custom CNN, and Xception net are considered for the proposed problem.

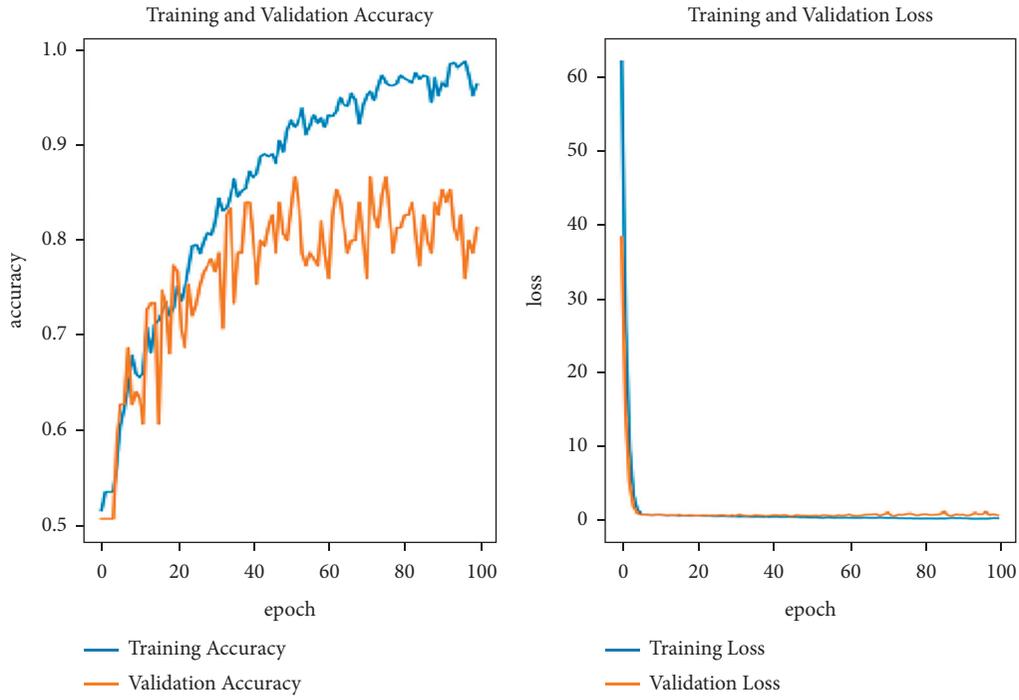


FIGURE 9: Accuracy and loss with Xception model.

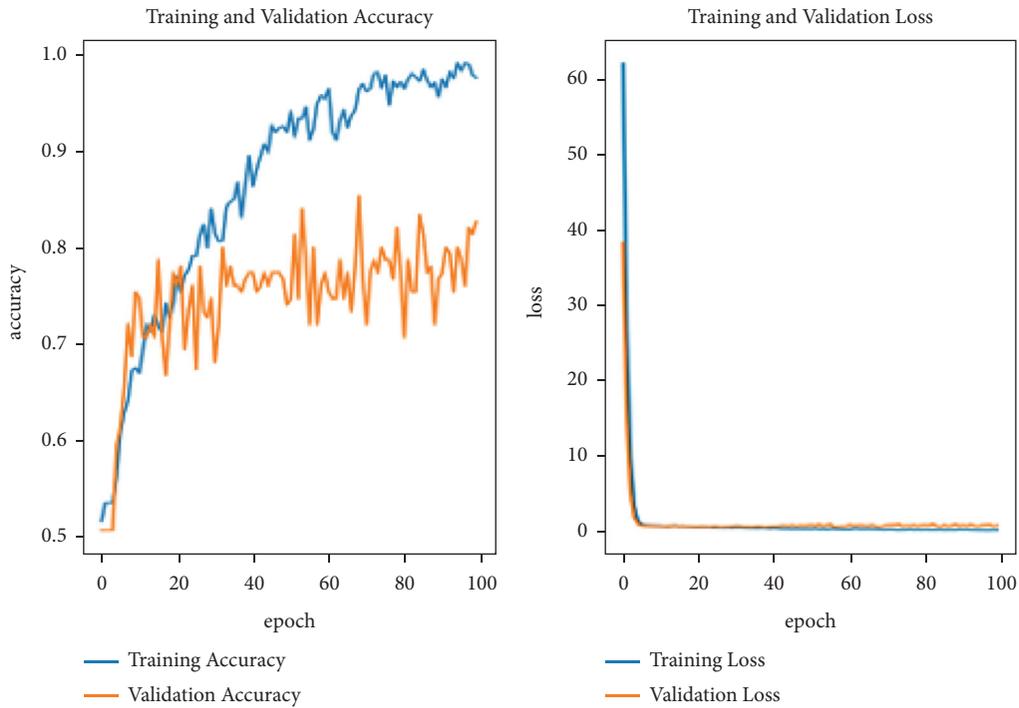


FIGURE 10: Accuracy and loss with AlexNet-based CNN model.

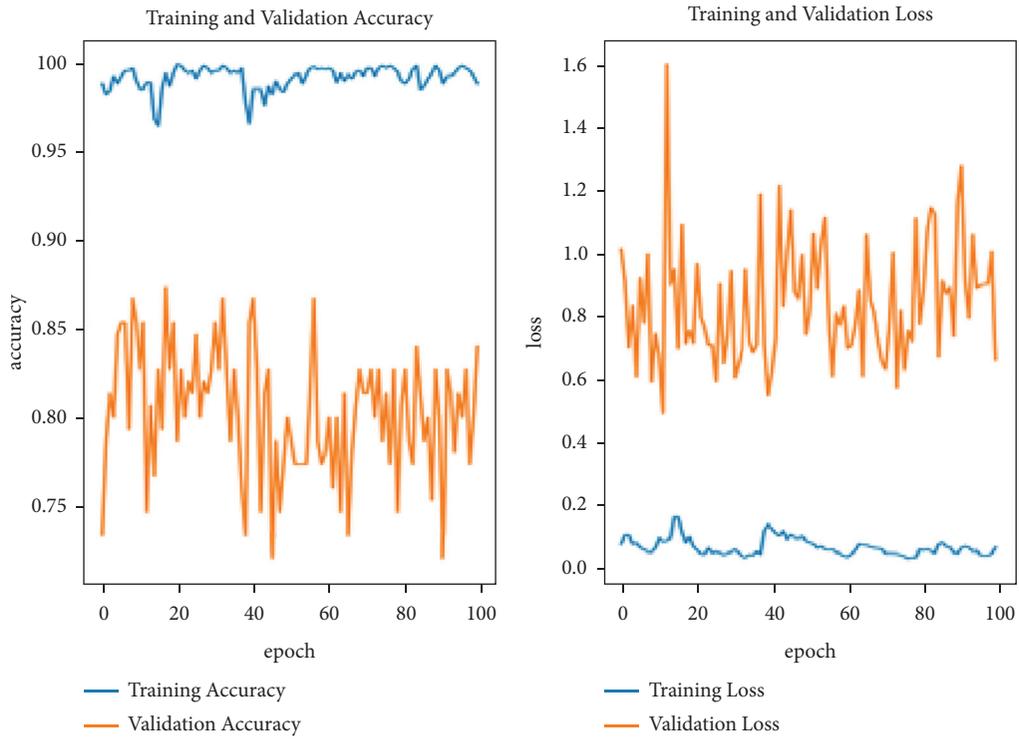


FIGURE 11: Accuracy and loss with ensemble transfer learning model.

TABLE 4: Performance analysis proposed method with the different models.

Ref	Classification types	Methods	Accuracy	Precision	Recall
[32]	Pothole classification	ResNet18	90.5	87.2	80.2
[33]	Pothole classification	Image processing	73.5	80.0	73.3
[34]	Pothole classification	LS-SVM	85.2	72.7	76.3
[34]	Pothole classification	ANN	88.7	85.7	78.2
[23]	Road damage detection	CNN	81.4	82.9	81.1
Present study	Bridge and road damage (5 classes)	Custom CNN	86.0	78.0	94.0
Present study	Bridge and road damage (5 classes)	Xception	84.0	79.0	89.0
Present study	Bridge and road damage (5 classes)	AlexNet	86.0	78.0	94.0
Present study	Bridge and road damage (5 classes)	Ensemble model	87.1	84.92	83.53

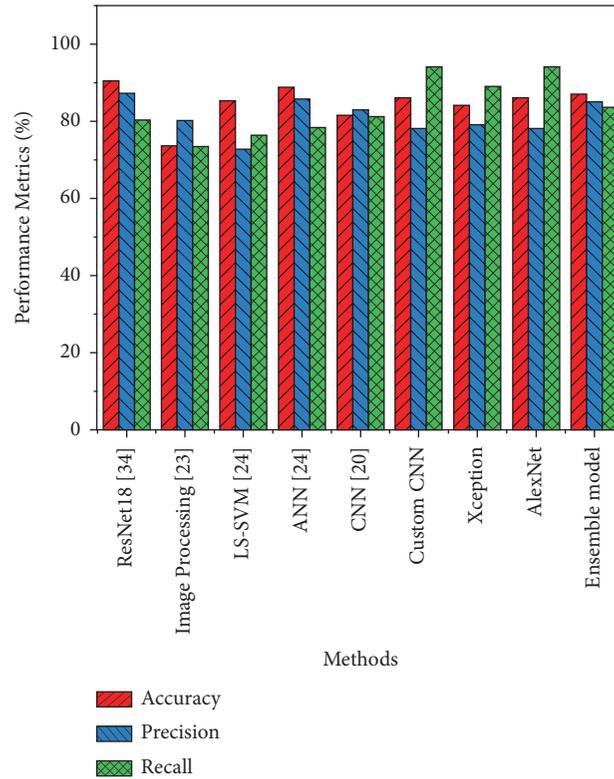


FIGURE 12: Performance analysis of proposed with different models.

TABLE 5: Run time for different pretrained and custom CNN models.

Model	Training time (sec/epoch)
AlexNet	13
VGG16	280
Xception	146
ResNet50	353
ResNet101	372
Custom CNN	24
Ensemble learning	216

6. Conclusion and Future Scope

Maintaining the civil infrastructure in a country like India is a challenging task. It is time-consuming and requires a lot of skilled workforces to identify the damages. In this regard, it is very significant to introduce a system to analyze the damages and detect them automatically based on vision-based techniques. Although successful strategies are available for road damage detection, they were not specific to Indian roads. In this proposed work, other civil infrastructure damages like bridge damages were also considered. The pretrained CNN architectures like Xception and AlexNet were used in the proposed work for transfer learning. To obtain more knowledge from each sample, the training data was expanded by the data augmentation method. At the end of this study, the ensemble transfer learning model's accuracy value was measured as 87.1% on the test data not seen before by the model. It is predicted that

higher accuracy can be achieved by adding more and more unique samples to the train data.

In the future, the dataset will be augmented using advanced techniques like undersampling and oversampling to reduce the dataset imbalance problem. Further, the proposed model will be deployed in the unmanned aerial vehicle to monitor the road damage condition and bridge condition during the natural disaster without risking human life.

Data Availability

The data used to support the findings of this study are included in the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- [1] K. Lee, N. Byun, and D. H. Shin, "A damage localization approach for rahmen bridge based on convolutional neural network," *KSCE Journal of Civil Engineering*, vol. 24, no. 1, pp. 1–9, 2020.
- [2] W. Wang, B. Wu, S. Yang, and Z. Wang, "Road damage detection and classification with faster R-CNN," in *Proceedings of the 2018 IEEE International Conference on Big Data (Big Data)*, pp. 5220–5223, SAN FRANCISCO, CA, USA, July 2018.
- [3] J. D. Sitton, Y. Zeinali, and B. A. Story, "Design and field implementation of an impact detection system using

- committees of neural networks,” *Expert Systems with Applications*, vol. 120, pp. 185–196, 2019.
- [4] N. Kumar, V. Poonia, B. B. Gupta, and M. K. Goyal, “A novel framework for risk assessment and resilience of critical infrastructure towards climate change,” *Technological Forecasting and Social Change*, vol. 165, no. January, Article ID 120532, 2021.
- [5] V. Hoskere, Y. Narazaki, T. A. Hoang, and B. F. Spencer, “Vision-based structural inspection using multiscale deep convolutional neural networks,” in *Proceedings of the 3rd Huixian International Forum on Earthquake Engineering for Young Researchers*, University of Illinois, Urbana-Champaign, US, August 2017.
- [6] Z. Zhang, C. Sun, R. Bridgelall, and M. Sun, “Road profile reconstruction using connected vehicle responses and wavelet analysis,” *Journal of Terramechanics*, vol. 80, no. November, pp. 21–30, 2018.
- [7] C. Yu, J. Li, X. Li, X. Ren, and B. B. Gupta, “Four-image encryption scheme based on quaternion Fresnel transform, chaos and computer generated hologram,” *Multimedia Tools and Applications*, vol. 77, no. 4, pp. 4585–4608, 2018.
- [8] Y. Qin, C. Xiang, Z. Wang, and M. Dong, “Road excitation classification for semi-active suspension system based on system response,” *Journal of Vibration and Control*, vol. 24, no. 13, pp. 2732–2748, 2018.
- [9] D. Lee, J. R. A. Felix, S. He, D. Offenhuber, and C. Ratti, “CityEye: real-time visual dashboard for managing urban services and citizen feedback loops,” in *Proceedings of the CUPUM 2015 - 14th Int. Conf. Comput. Urban Plan. Urban Manag.*, Cambridge, Massachusetts USA, July 2015.
- [10] D. Li, L. Deng, B. Bhooshan Gupta, H. Wang, and C. Choi, “A novel CNN based security guaranteed image watermarking generation scenario for smart city applications,” *Information Sciences*, vol. 479, pp. 432–447, 2019.
- [11] S. R. Sahoo and B. B. Gupta, “Multiple features based approach for automatic fake news detection on social networks using deep learning,” *Applied Soft Computing*, vol. 100, Article ID 106983, 2021.
- [12] B. F. Spencer, “Impact of digital fingerprint image quality,” *Engineering*, vol. 5, no. 2, pp. 1–40, Jan. 2019.
- [13] V. Chandran, C. K. Patil, A. Karthick, D. Ganeshaperumal, R. Rahim, and A. Ghosh, “State of charge estimation of lithium-ion battery for electric vehicles using machine learning algorithms,” *World Electric Vehicle Journal*, vol. 12, no. 1, p. 38, 2021.
- [14] J. J. Lee, J. W. Lee, J. H. Yi, C. B. Yun, and H. Y. Jung, “Neural networks-based damage detection for bridges considering errors in baseline finite element models,” *Journal of Sound and Vibration*, vol. 280, no. 3–5, pp. 555–578, 2005.
- [15] M. Mehrjoo, N. Khaji, H. Moharrami, and A. Bahreininejad, “Damage detection of truss bridge joints using Artificial Neural Networks,” *Expert Systems with Applications*, vol. 35, no. 3, pp. 1122–1131, 2008.
- [16] H. M. Ngwangwa, P. S. Heyns, F. J. J. Labuschagne, and G. K. Kululanga, “Reconstruction of road defects and road roughness classification using vehicle responses with artificial neural networks simulation,” *Journal of Terramechanics*, vol. 47, no. 2, pp. 97–111, 2010.
- [17] X. Yu and E. Salari, “Pavement pothole detection and severity measurement using laser imaging,” in *Proceedings of the 2011 IEEE International Conference on Electro/Information Technology*, MI, USA, May 2011.
- [18] L. Li, L. Sun, G. Ning, and S. Tan, “Automatic pavement crack recognition based on BP neural network,” *Promet - Traffic & Transportation*, vol. 26, no. 1, pp. 11–22, 2014.
- [19] T. Lee, C. Chun, and S.-K. Ryu, “Detection of road-surface anomalies using a smartphone camera and accelerometer,” *Sensors*, vol. 21, no. 2, p. 561, 2021.
- [20] S. Li, Y. Cao, and H. Cai, “Automatic pavement-crack detection and segmentation based on steerable matched filtering and an active contour model,” *Journal of Computing in Civil Engineering*, vol. 31, no. 5, Article ID 04017045, 2017.
- [21] F. Sarie, M. Bisri, A. Wicaksono, and R. Effendi, “Types of road pavement damage for road on peatland, A study case in palangka raya, central kalimantan, Indonesia,” *IOSR J. Environ. Sci*, vol. 9, no. 12, pp. 53–59, 2015.
- [22] H. Maeda, Y. Sekimoto, and T. Seto, “Lightweight road manager,” in *Proceedings of the 5th ACM SIGSPATIAL International Workshop on Mobile Geographic Information Systems*, pp. 37–45, Burlingame, California, October 2016.
- [23] Y.-J. Cha, W. Choi, and O. Büyükköztürk, “Deep learning-based crack damage detection using convolutional neural networks,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 5, pp. 361–378, 2017.
- [24] C. Kyriakou, S. E. Christodoulou, and L. Dimitriou, “Smartphone-based pothole detection utilizing artificial neural networks,” *Journal of Infrastructure Systems*, vol. 25, no. 3, Article ID 04019019, 2019.
- [25] A. Karmel, M. Adhithyan, and P. Senthil Kumar, “Machine learning based approach for pothole detection,” *International Journal of Civil Engineering & Technology*, vol. 9, no. 5, pp. 882–888, 2018.
- [26] Q. Zou, Z. Zhang, Q. Li, X. Qi, Q. Wang, and S. Wang, “DeepCrack: learning hierarchical convolutional features for crack detection,” *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1498–1512, 2019.
- [27] Y. Jo and S. Ryu, “Pothole detection system using a black-box camera,” *Sensors*, vol. 15, no. 11, pp. 29316–29331, 2015.
- [28] A. Cubero-Fernandez, F. J. Rodriguez-Lozano, R. Villatoro, J. Olivares, and J. M. Palomares, “Efficient pavement crack detection and classification,” *EURASIP Journal on Image and Video Processing*, vol. 2017, no. 1, 2017.
- [29] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, and H. Omata, “Road damage detection and classification using deep neural networks with smartphone images,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 12, pp. 1127–1141, 2018.
- [30] J. Lekshmiopathy, N. M. Samuel, and S. Velayudhan, “Vibration vs. vision: best approach for automated pavement distress detection,” *International Journal of Pavement Research and Technology*, vol. 13, no. 4, pp. 402–410, 2020.
- [31] M. Suriya, V. Chandran, and M. G. Sumithra, “Enhanced deep convolutional neural network for malarial parasite classification,” *International Journal of Computers and Applications*, pp. 1–10, 2019.
- [32] Aparna, Y. Bhatia, R. Rai, V. Gupta, N. Aggarwal, and A. Akula, “Convolutional neural networks based potholes detection using thermal imaging,” *Journal of King Saud University - Computer and Information Sciences*, 2019.
- [33] N.-D. Hoang, “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*, vol. 2018, pp. 1–12, 2018.
- [34] S.-K. Ryu, T. Kim, and Y.-R. Kim, “Image-based pothole detection system for ITS service and road management system,” *Mathematical Problems in Engineering*, vol. 2015, pp. 1–10, 2015.