Classification of Road Features Using Transfer Learning and Convolutional Neural Network (CNN)

Ahmed Abed Mohammed[1,2] **and Putra Sumari**[3]

[1]*Department of Technical Engineering, Islamic University, Najaf, Iraq*
[2]*Department of Technical Engineering, Islamic University, Dewania, Iraq*
[3]*School of Computer Science, University Sains Malaysia, 11800 Peneng, Malaysia*

Receive Mon. 20, Revise Mo. 20, Accepte Mon. 20, Publishe Mon. 20

Abstract: Efficient and accurate classification of road features, such as crosswalks, intersections, overpasses, and roundabouts, is crucial for enhancing road safety and optimizing traffic management. In this study, we propose a classification approach that utilizes the power of transfer learning and convolutional neural networks (CNNs) to address the road feature classification problem. By leveraging advancements in deep learning and employing state-of-the-art CNN architectures, the proposed system aims to achieve robust and real-time classification of road features. The dataset contained 7616 images of roundabouts, crosswalks, overpasses, and intersections from the MLRSNet dataset and manually extracted satellite images from Malaysia using Google Earth Pro. After that, we merged this dataset. We designed a CNN architecture that consists of 24 convolution layers and eight fully connected layers. Transfer learning models such as ResNet50, MobileNetV2, VGG19 and InceptionV3 were also explored for road feature classification. The best-performing model during the validation phase is InceptionV3, with an accuracy of 98.9777%, whereas the best-performing model during the test phase is ResNet50 and VGG-19 models, with an accuracy of 98.7132%. The proposed CNN model got 95.1208% and 94.4852% accuracy during the validation and test stage. From the evaluation, the best-performing models for road feature classification are ResNet50 and VGG-19, with an accuracy of 98.7132%.
Keywords: Deep learning, Transfer Learning, Road feature, Classification.

1. INTRODUCTION

The speedy growth of computer vision and deep learning has been a crucial factor in the automated extraction and classification of road parameters. Road feature classification is key to many applications like GIS, autonomous driving, and urban planning. Merging of Transfer Learning and Convolutional Neural Networks (CNNs) is a sound method to boost the speed and accuracy of road feature classification. Effective classification of road features is essential for applications in transportation systems, such as autonomous driving, traffic management, and road infrastructure maintenance[1]; for safe and effective navigation, it is crucial to correctly identify and classify road elements such as roundabouts, crosswalks, overpasses, and crossroads[2], road feature classification traditionally depends on labour-intensive, subjective, and error-prone manual examination and interpretation by human experts[3].Automated categorization systems, however, have emerged as a potential approach to handle this work effectively and precisely thanks to improvements in computer vision, machine learning, and deep learning approaches[4]; various strategies have been investigated in recent years to automate the classification of road features[5], These methods utilize computer vision techniques and machine learning models, particularly convolutional neural networks (CNNs), to analyze and categorize photographs of road features. With the ability to extract pertinent information from input photos and produce precise predictions, CNNs have shown to be exceptionally effective at image recognition tasks. In this paper, we address the classification issue of these unique road features by developing a convolutional neural network (CNN) architecture and using a transfer learning technique to search for the best-performing model[6]. CNNs prove to be able to independently and continuously acquire feature hierarchies from input images. As a result, they are components of deep neural networks that have exceptional performance in image processing applications. Besides, computer vision applications like image segmentation and picture classification have also proved to be very efficient and successful in operating using Convolutional Neural Networks (CNNs). CNNs can recognize various road types, road markings, signs, and others by analyzing high-resolution remote sensing images or ground-level photographs. This ability is of great use in the classification of road attributes.Transfer Learning is a machine learning method that involves the use of a pre-trained model to create a new model for a different task.

Through this approach, you can improve your performance on the current task by taking the knowledge and skills from the previous one and applying them. Transfer learning allows the employment of complex models that have already been trained on big datasets like ImageNet to identify the basic characteristics from the road images, thus simplifying the process of classifying the road features. In specific fields like road feature recognition, the availability of large labelled datasets is usually limited. Thus, the importance of computer vision algorithms is diminished. Through the combination of CNNs with transfer learning, the most beneficial features of the two methods may be captured. On the other hand, with a minimal quantity of road-specific data, the pre-trained CNN models can be easily fine-tuned to recognize and classify the elements of the road with great accuracy. In this way, we will be able to speed up the training, and the model's generalizability will be improved. Hence, it will be more effective when used for new data that have not been known before.A lot of the research has proved that this approach is successful in providing the best solutions. For instance, scientists have accomplished a lot in the segmentation of road types (like highways, urban streets and country roads) and road conditions (like wet, dry and ice) from aerial and ground-level images by using transfer learning. Besides, to enhance the classification system's robustness and accuracy, CNNs can be trained to use multi-modal input, for instance, to combine visual pictures and LiDAR data. In a nutshell, the best way to categorize road features that make progress in data scarcity and demand high accuracy is the combination of transfer learning and convolutional neural networks (CNNs). Intelligent transport systems and smart city applications alike benefit from this holistic strategy, in which the road features are detected and classified more efficiently and effectively. The advancements in the road feature categorization accuracy and sophistication, which are going to be achieved through research and development in this field, will be in line with the ongoing technological transformation, meaning that the new technologies will bring huge benefits to the field. The successful implementation of this research will have significant practical implications. It can assist traffic management authorities in automating the monitoring and controlling of transportation facilities, enhancing road safety, and optimizing traffic flow. Additionally, the proposed system can serve as a foundation for developing intelligent transportation systems and smart city initiatives, contributing to the overall improvement of urban mobility.

2. Related Works

In recent years, there has been much interest in the classification of road characteristics, including roundabouts, crosswalks, overpasses, and intersections, utilizing transfer learning and convolutional neural networks (CNNs). Researchers have investigated several methods to improve the reliability and accuracy of classifying road features in the context of intelligent transportation systems.

Tümen et al. proposed deep learning and image-processing techniques to detect intersections and crosswalks. They designed a multi-scale CNN architecture called the RoIC-CNN that incorporated both convolutional and pooling layers to capture spatial information at different scales. RoIC-CNN consists of ten convolution layers and eight fully connected layers. In this study, other CNN models, such as VggNet-5, LeNet, and AlexNet, are also tested to compare their performance. From the evaluation, the best-performing model in detecting crosswalks and intersections is the RoIC-CNN[7]; In order to create the flyover labelling geodatabase (OLGDB) using the OpenStreetMap (OSM) road network data of six representative Chinese cities, Li et al. build upon the target detection model (Faster-RCNN). This method uses raster data to train convolutional neural networks (CNNs) to learn task-adaptive features, and then uses a region proposal network (RPN) to choose the best spot for a flypast. More specifically, Faster-RCNN incorporates three separate CNNs: ZF-net, VGG-16, and Inception-ResNet V2. The contribution of five geometric metrics—area, perimeter, squareness, circularity, and W/L—to the flyover identification task is assessed after synthesising them into picture bands to improve the training data. In this stage, the optimal combination of learning rate and batch size is determined by fine-tuning. The suggested method achieves respectable accuracy (about 90%) according to the experimental findings[8]. Another active area of research aimed at leveraging artificial intelligence techniques to improve transportation infrastructure and operations is the analysis of the deep learning-based classification of transportation facilities for enhanced road safety and traffic management. In the study conducted by Jilani et al. (2022), When it comes to traffic congestion categorization, a five-layer CNN deep learning model is suggested. Augmentation with GANs improves the traffic congestion dataset. The study used pre-trained RsNet50 and DenseNet-121 as the benchmark to compare with the 5-layer CNN. The study found that the proposed CNN emerged as the best model with an accuracy of 98.63% compared to ResNet50 (90.59%) and DenseNet-121 (93.15%), respectively[9]; This study presents a hybrid model that enhances road feature identification by merging Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). The main objective is to enhance road safety and optimize the overall driving experience. The model employs Convolutional Neural Networks (CNNs) based on the MobileNetV2 architecture and Recurrent Neural Networks (RNNs) to handle GRU. A benefit of using it in automobiles is its ability to process real-time data and effortlessly transfer models due to its lightweight form. The CNN-MobileNetV2-GRU model has improved its ability to accurately identify road components, including speed bumps and variations in road conditions. The model provides a comprehensive solution for classifying road features, including spatial, efficiency, and speed requirements, by processing data from the edge to the cloud. The CNN-MobileNetV2-GRU model is well-suited for practical applications such as enhanced driver assistance systems and autonomous vehicles because of its exceptional precision and efficient utilization of processing resources. [10]; The authors propose a technique that classifies roads into four categories: highway, city road, undeveloped area, and housing estate. A sophisticated pyramidal residual network has been created to classify these sorts of roads effectively. Conducted experiments using a benchmark dataset that is accessible to the public and sensor data obtained via intelligent eyewear. The 1D-PyramidNet model had the highest level of accuracy (92.23%) in interpreting the data and surpassed all other deep learning models in performance[11]. As detecting road damage is important to maintain optimal road conditions and enhance transportation safety, several studies have been using object detection/classification models and incorporating a deep learning technique. Road damage algorithms can be divided into two large groups; the first group includes algorithms with two stages, and the second consists of one-stage ones. A two-state solution, like support vector machines and convolutional neural networks, includes a detector and a classifier, which first defines regions in an image where an object may be present and then classifies each area. Another type, a one-stage solution, for instance, YOLO and SSD, tries to make predictions for the best possible region at one time, negating the need for additional steps. This work describes the development of such algorithms and demonstrates their use in examples of explaining road damage detentions[12]. Deep learning algorithms are now being used in vision-based applications, such as autonomous driving and traffic monitoring. Traffic sign recognition and semantic road detection are the primary areas of study in intelligent transport systems, with a strong emphasis on safety. These issues play a crucial role in the advancement of intelligent transport systems. This work introduces a driving assistance system that incorporates components based on deep learning. The system is constructed using hybrid 2D-3D CNN models that use transfer learning techniques. The models use a pre-trained deep 2D Convolutional Neural Network (CNN) and a less complicated 3D CNN to reduce complexity and expedite the training method. The first model, Hybrid-TSR, is an established method for handling the problem of recognizing traffic signs. The second model, Hybrid-SRD, is a technique for detecting road space by using up-sampling and deconvolutional operations to analyze the semantic information. The projected results indicate that the offered approaches are very significant in terms of efficiency and accuracy[13]. Autonomous cars are significantly important for traffic moving monitoring, and the capability of instantly detecting potholes is important for the safety and convenience of the vehicles. Many methods, such as reporting to authorities, vibration-based sensors, and 3D laser imaging, are limited by the high costs that are incurred in their installation and the possible dangers that are associated with their use. This article presents the new method, Adaptive Mutation and Dipper Throated Optimisation (AMDTO), which is designed to select and optimize the features of the Random Forest (RF) classifier. The AMDTO+RF technique that was employed had a pothole classification accuracy of 99. The effectiveness of the method in experiment A was 795%, which was beyond the previous methods, such as WOA+RF, GWO+RF, PSO+RF, and transfer learning approaches. The method's importance and consistency are even more proven by the in-depth statistical analysis of the outcomes that are recorded. This method aims to electronify the process of spot identification that is precise and quick[14]. Automated data collecting for roadside barriers has been developed by the Wyoming Department of Transportation (WYDOT) as part of their asset management system. The system collects the geometric attributes and material conditions of barriers, which in turn assists in the decision-making of asset management and, thus, optimization. There are over one million linear feet of state barriers, whose total value is more than $100 million. The price for the state to acquire these features is more than half a million dollars at one time. A unique method was suggested to identify different kinds of roadside barriers by using pre-trained models like inception v3, denseness 121, and VGG 19. VGG 19 network was used, which resulted in a great accuracy of 97% through transfer learning. An architectural non-transfer model, which is basically a model that is just built and is very simple, was made and the accuracy of the model was 85%. On the other hand, the non-transfer learning model was better than the inception and denseness models but still not as good as the VGG network.[15]; The study introduces a technique for classifying road signs using pre-trained Convolutional Neural Network (CNN) models, which is based on transfer learning. The authors assess the efficacy of their models using the German Traffic Sign Recognition Benchmark test dataset. The researchers use transfer learning and augmentation approaches to assess different designs. The findings demonstrate that the suggested strategy attains an average accuracy of 99.2%, surpassing the performance of current approaches. These findings indicate that the use of transfer learning and pre-trained models may greatly improve the accuracy of road sign categorization, even when working with a limited dataset[16]. This research assesses the efficacy of deep learning-based pre-trained networks in assessing gravel road photographs using traditional methodologies. The collection comprises photos obtained from self-recorded films and Google Street View. These images have been manually tagged based on standard images established by the Road Maintenance Agency in Sweden. The dataset was divided into a 60:40 ratio for training and testing. Multiple pre-trained models were used, all of which exhibited strong performance with an accuracy of over 92%. The VGG-16 model, which was pre-trained and used transfer learning, had superior performance in terms of accuracy and F1 score when compared to other models that were suggested. The research seeks to enhance the evaluation of loose gravel by road maintenance authorities[17]. There is an increasing need to identify wet road surfaces in order to address accidents and traffic problems during rainy weather conditions. Acoustic signals have garnered interest because of their cost-effectiveness in deployment.A large quantity of training data is required by current deep learning methods, which rely on supervised audio measurements. The evolution of convolutional neural networks (CNNs) has made it easier to train CNNs on one dataset and then apply them to another. The reliability of convolutional neural network (CNN) models that have been pre-trained to detect wet road surfaces is tested in this study. The results show that transfer learning is able to distinguish between surfaces that are dry and those that are moist, with an accuracy rate of more than 80%.[18].

3. Methodology

The methodology consists of a set of phases: dataset collection, preprocessing, Classification Models, Performance Metric and Hyperparameter Tuning, as shown in Figure 1.

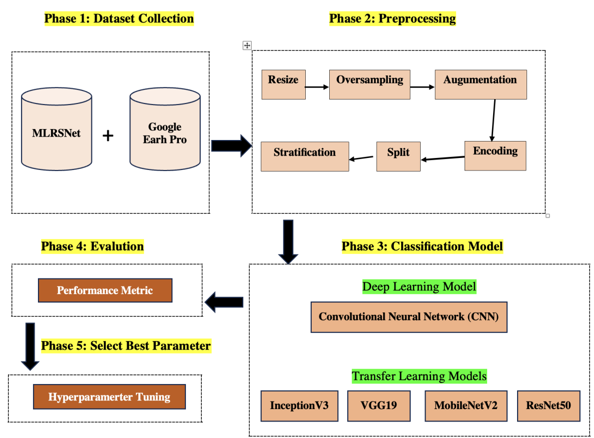

Figure 1. Methodology of the study.

A. Phase 1:Dataset Description

We will explain the dataset used in our study.

1- MLRSNet Dataset (First Dataset): The dataset used in this study is primarily sourced from the MLRSNet dataset, which is available from Mendeley Data [19]. The dataset contains 109,621 high spatial resolution optical images of 46 different categories captured from satellites. In this study, images of roundabouts, intersectionsIntersection and overpasses were obtained from the MLRSNet dataset. 2,040, 2,498 and 2500 images of roundabouts, Intersections and overpasses were collected from the MLRSNet dataset, respectively. The images from the MLRSNet dataset have a fixed size of 256x256 pixels, as show in Figure 2.

2- Dataset from Google Earth Pro (Second Dataset): is provides users with access to a wealth of geographical data, including satellite photos, maps, topography, and 3D buildings, all from the comfort of their virtual globe [20]. Additional satellite images of roundabouts, intersections, crosswalks, and overpasses in Malaysia were obtained by taking screenshots of those features using Google Earth Pro software that contain 100, 101, 242 and 135 Roundabout, Intersections, Crosswalk and Overpass as show in Figure 3.

3- Merge Dataset: We merged two datasets, the MLRSNet dataset and the dataset from Google Earth Pro, to

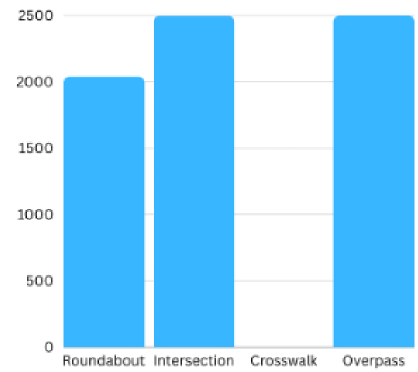

Figure 2. MLRSNet Dataset

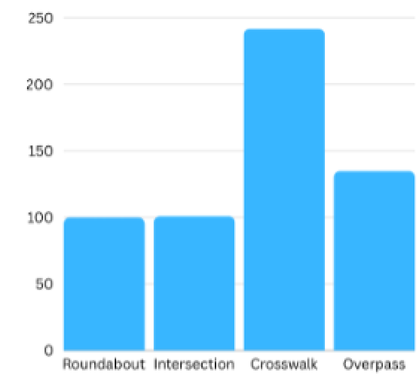

Figure 3. Dataset from Google Earth Pro

get 2140, 2599, 242 and 2635 to Roundabout, Intersection, Crosswalk and Overpass as Figure 4. and Table I, As can be observed from Table I and Figure 4., the data could be more balanced, where the number of images of crosswalks is significantly less than the images of other categories since the images of crosswalks are not explicitly available to be downloaded from the MLRSNet Dataset. Figure 5 shows the sample images for the roundabout, Intersection, crosswalk, and overpass.

B. Phase 2: Preprocessing

We apply a set of processes before entering the classification model

1- Resizing Image: It is essential preprocessing to change the size of the Image and its dimensions while maintaining its aspect ratio or stretching it to fit a new size. To ensure uniformity and compatibility, all photos in our dataset were uniformly downsized to a resolution of 224x224 pixels. The raw images collected from Google

TABLE I. Number of Image to each Class

Images	MLRSNet	Google Earth Pro	Total
Roundabout	2040	100	2140
Intersection	2498	101	2599
Crosswalk	0	242	242
Overpass	2500	135	2635

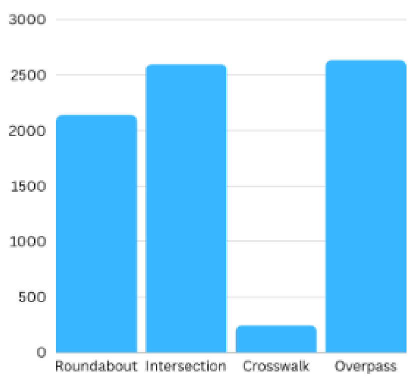

Figure 4. Merge Dataset

Figure 5. Sample of Images

Earth Pro were of a different size and aspect ratio than the MLRSNet dataset. Thus, the images obtained from Google Earth Pro were cropped into perfect squares. This was realized through checking the aspect ratio of the images. If the image width and height are equal, no cropping action will be performed. If the width is greater than the height, the left and right sides of the Image will be cropped, and a perfect square image is returned. If the width is less than the height, the top and bottom parts of the Image will be cropped, and a square image will be returned. The cropped images will then be resized to a resolution of 224×224. Similarly, images obtained from the MLRSNet dataset were resized to 224x224 pixels as well.

2- Oversampling: is a method for dealing with datasets that have an imbalance between classes, where one class has a much smaller number of instances than the other classes. Biassed models that fail to represent minority groups adequately may result from this imbalance; class imbalance is detrimental towards training on classifiers, and it affects the convergence of the deep learning model during the training phase and generalization of the model during the test phase [21], [22]. Good model results can be attained if all the classes in the classifier are properly represented [23]. The oversampling method was found to be among the best methods in alleviating class imbalance problems for CNN-related model training [21]. In order to alleviate the class imbalance issue in this study, image augmentation, which is a form of the oversampling method, was performed so that class balance was achieved. The argumentation library from Python was used to perform image augmentation. In this data augmentation process, for each image class, a random image was selected, and random augmentation operations were performed. The following augmentation techniques, with a 0.5 probability chance of execution, were performed in Figure 6

- Random rotate- 90°
- Vertical flip
- Horizontal flip
- Random brightness contrast
- Random gamma

3- Image Augmentation: in order to increase the size of the dataset, we used methods for augmenting images. Incorporating data variances via image augmentation is a common way to improve the general-isolation Model's performance[24]. Through data augmentation, the number of images for each category is increased to 2700 images, respectively, as Figure 7.

4- Encoding: is a method to change the way values are represented. When dealing with categorical variables, label encoding—also called ordinal Encoding gives each category in the dataset a distinct integer value [25]. The

Figure 6. Some of Edits on Image

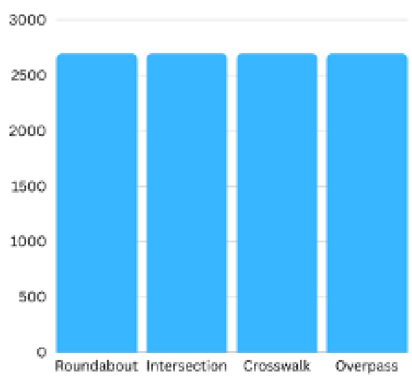

Figure 7. Number of Image after Image Augumentation

input and filter weights are determined by doing the dot product using this technique, which enables the conversion of convolutional filters. The network is able to analyse the input image and extract relevant features and spatial information by means of these procedures [26].

5- Split Data: The images are then split to train, validate, and test set with a ratio of 7:2:1. Here we have the train set, which is used to train the model; the validation set, which is used to validate the model after each epoch; and the test set, which is used to assess the model after training.

6- Stratification: is a method for making sure that various data subsets keep the same distribution of classes in the dataset. Because biassed models might result from an uneven class distribution, this is especially crucial for classification tasks, which were performed during the split to ensure that the number of images for every class was the same for each batch of dataset.

C. Phase 3: Classification Models

1- Convolutional Neural Network (CNN): Among artificial neural networks, Convolutional Neural Networks (CNNs) stand out due to their deep feed-forward design [27]. Image processing and analysis are two areas where Convolutional Neural Networks (CNNs) really shine. In order to assess the input picture, a Convolutional Neural Network (CNN) usually uses convolutional layers that use sets of adaptive filters. CNN is well-known in image-based classification tasks. CNNs process the input data using a number of interconnected layers. Typically, a convolutional layer is the first hidden layer of a convolutional neural network (CNN). It uses a series of filters to identify patterns in the input data [28]. These networks are built to automatically extract significant features at various degrees of abstraction from raw pixel input and learn hierarchical representations. Convolutional layers in CNNs allow for the extraction of regional patterns and structures. In contrast, pooling layers make it easier to downsample spatial data, which improves the model's ability to identify essential features. Based on the retrieved features, fully connected layers provide the final categorization.

A custom CNN model is created in this study to perform the classification of roundabouts, crosswalks, intersections, and overpasses. The CNN model is created using TensorFlow. The CNN architecture consists of 24 convolution layers and eight fully connected layers, as listed in Table II.

3- Transfer Lerning: In addition to the proposed CNN, transfer learning models such as ResNet50, MobileNetV2, VGG19 and InceptionV3 are used to compare the accuracy with our CNN model developed. It is a powerful technique in computer vision tasks which enables models to leverage knowledge learned from pre-trained models trained on large-scale datasets [29]. By transferring this knowledge to a new task, transfer learning can significantly enhance classification performance. In our approach, we harness the benefits of transfer learning by utilizing pre-trained models from the literature, such as ResNet50, MobileNetV2, VGG19 and InceptionV3. In this study, the classification layers of the transfer learning models were dropped and replaced with a new classification layer similar to the proposed CNN model. The weights are initialized based on the models' weights trained on ImageNet. All the transfer learning layers were allowed to be trained in this study. Table III. summarizes the common hyperparameters configured for the transfer learning models.

- ResNet50: is a 50-layer deep convolutional neural network design that uses residual connections to support the

TABLE II. The CNN architecture

Layer	Layer Name	Output Shape	Param #
1	Conv2D	224 x 224 x 32	896
2	MaxPooling2D	112 x 112 x 32	0
3	Conv2D	112 x 112 x 32	9248
4	BatchNormalization	112 x 112 x 32	128
5	MaxPooling2D	56 x 56 x 32	0
6	Dropout	56 x 56 x 32	0
7	Conv2D	56 x 56 x 64	18496
8	BatchNormalization	56 x 56 x 64	256
9	Conv2D	56 x 56 x 64	36928
10	BatchNormalization	56 x 56 x 64	256
11	MaxPooling2D	28 x 28 x 64	0
12	Dropout	28 x 28 x 64	0
13	Conv2D	28 x 28 x 128	73856
14	BatchNormalization	28 x 28 x 128	512
15	Conv2D	28 x 28 x 128	147584
16	BatchNormalization	28 x 28 x 128	512
17	MaxPooling2D	14 x 14 x 128	0
18	Dropout	14 x 14 x 128	0
19	Conv2D	14 x 14 x 256	295168
20	BatchNormalization	14 x 14 x 256	1024
21	Conv2D	14 x 14 x 256	590080
22	BatchNormalization	14 x 14 x 256	1024
23	MaxPooling2D	7 x 7 x 256	0
24	Dropout	7 x 7 x 256	0
25	Flatten	12544	0
26	Dense	256	3211520
27	BatchNormalization	256	1024
28	Dropout	256	0
29	Dense	32	8224
30	BatchNormalization	32	128
31	Dropout	32	0
32	Dense	4	132

TABLE III. The common hyperparameters configured for the transfer learning

Parameter	InceptionV3	ResNet50	VGG-19	MobileNetV2
Input image	(224,224,3)	(224,224,3)	(224,224,3)	(224,224,3)
Weight	Initialized to ImageNet	Initialized to ImageNet	Initialized to ImageNet	Initialized to ImageNet
Optimizer	Adam	Adam	Adam	Adam
Loss function	Sparse categorical cross entropy	Sparse categorical cross entropy	Sparse categorical cross entropy	Sparse categorical cross entropy
Classifier	Softmax	Softmax	Softmax	Softmax
Epochs	50	50	50	20
Dropout rate	0.2	0.2	0.2	0.2

training of intense networks. It has demonstrated impressive performance in image classification challenges, overcoming the vanishing gradient issue and facilitating more straightforward deep model optimization.

- MobileNetV2: A compact convolutional neural network architecture called MobileNetV2 was created for effective mobile and embedded vision applications. It is appropriate for devices with limited resources because it uses depth-wise separable convolutions and inverted residual blocks to simplify computations while retaining competitive accuracy.

- VGG19: Convolutional neural network architecture VGG19 is renowned for its efficiency and simplicity. There are 19 layers total, including several 3×3 convolutional layers followed by max-pooling layers. Although VGG19 has more parameters than other architectures, it performs image classification tasks with a high degree of accuracy.

- InceptionV3: The deep convolutional neural network architecture known as InceptionV3 makes use of the idea of inception modules. These modules use parallel convolutional layers with various kernel sizes to capture features at various scales. InceptionV3 reduces the number of parameters using dimensionality reduction techniques to obtain high accuracy in image recognition tasks while retaining computing efficiency.

D. Phase 4: Performance Metric

In this stage, we test the CNN models to see how well they do. To gauge how well a Convolutional Neural Network (CNN) is doing, accuracy is a common metric to utilise. Relative to the total number of photos, accuracy measures how many predictions were right. Equation 1 is the accuracy formula:

$$Acc(\%) = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \qquad (1)$$

True positives (TPs) occur when the model accurately predicts that a sample belongs to a given class, while true negatives (TNs) occur when the model accurately predicts that a sample does not belong to a specific class. A False Positive (FP) occurs when the model wrongly assigns a sample to the wrong class, whereas a False Negative (FN) occurs when the model assigns the wrong class to a sample, even if the sample truly belongs to the proper class.

E. Phase 5: Hyperparameter Tuning

The optimal model's performance was achieved by hyperparameter adjustment. Two hyperparameters—learning rate and batch size—were examined in this research. According to [30], the learning rate is a hyperparameter that controls how much a deep learning model changes whenever the model's weights are changed in response to the estimated error. Batch size, on the other hand, is the quantity of samples that are processed prior to updating the model [31]. Learning rate and batch size significantly affect the performance of a neural network, according to several research [32][33].

Learning rates of 0.01, 0.001 and 0.0001 were tested during hyperparameter tuning. As for batch size, batch sizes of 32 and 64 were tested. A full grid search was performed during the hyperparameter tuning process, and thus, six experiments were performed for each of the image classification models during the hyperparameter tuning process.

The hyperparameter tuning process is realized using the Keras Tuner library in Python, and the models were allowed to be trained for 50 epochs. After hyperparameter tuning was performed for each image classification model, the model with the best validation accuracy was rebuilt, and the model was then tested with the test dataset to obtain the test accuracy.

Subsequently, the performance of the best-performing model for each image classification model, whether it is the custom CNN model proposed here or the transfer learning model, will be evaluated, compared and discussed in a subsequent section.

4. Results and Discussion

We will Split this chapter to two section Result and Discussion

A. Results

We will show all results that we achieved in our study

1- Training and Validation Loss: Figure 8 shows the training and validation loss of the model during the hyperparameter tuning process. In general, the training and validation loss reduces as the epoch increases. However, some of the trials failed to converge and remain stationary across epochs, which is especially true for trial 1 and trial 2 of transfer learning models, where the learning rate for both was 0.01. While trial 5 and trial 6, which use a learning rate of 0.0001, show the lowest training and validation loss during the model training of transfer learning models, both trials had the highest loss during the training of the proposed CNN networks.

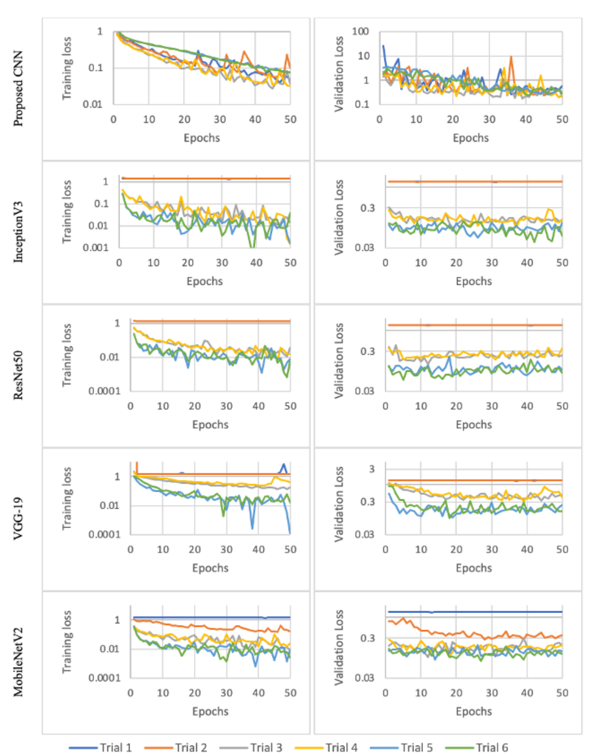

Figure 8. Training and validation loss of models plotted on log scale during the hyperparameter tuning process.

2- Validation Accuracy: Table IV summarizes the proposed CNN and transfer learning models' best validation accuracy for different learning rates and batch sizes used during the hyperparameter tuning process. In general, in this study, a larger batch size of 64 would result in better validation accuracy than a smaller batch size of 32; as for the effects of learning rate, smaller learning rates result in better validation accuracy for transfer learning models. However, this trend is not observed in the proposed CNN model, where the highest validation accuracy is observed when the learning rate is 0.001 instead.

It is worth noting that for all the transfer learning models trained with a learning rate of 0.01, except for the MobileNetV2 model trained with a batch size of 64, the accuracies of the models were very low at less than 40%. This was due to the models failing to converge during model training, as evident in the training and validation loss curves of those models shown in Figure 8, which do not decrease as the training epochs increase.

3- Optimal Hyperparameters: The highest validation accuracies for each model in the hyperparameter tuning process were bolded in Table V, and accuracy values, together with their corresponding hyperparameters, were summarized in Table 5. From the hyperparameter tuning process, the best-performing proposed CNN model, which had a learning rate of 0.001 and batch size of 32, had a validation accuracy of 95.1208%. As for other transfer learning models, the highest validation accuracies were attained with a learning rate of 0.0001 and batch size of 64. The validation accuracies were 98.9777%, 98.6524%, 97.9082% and 98.3271% for InceptionV3, ResNet50, VGG-19 and MobileNetV2 respectively. Based on this validation accuracy score, InceptionV3 was found to provide the best validation accuracy, followed by ResNet50, MobileNetV2, VGG-19 and the proposed CNN network.

In addition, the best-performing models were tested with the test dataset to obtain the test accuracy. In general, the test accuracy was comparable to the validation accuracy. However, the sequence in terms of the test accuracy performance is different. Both ResNet50 and VGG-19 had a test accuracy of 98.7132%, followed by MobileNetV2 with 98.0698% test accuracy and InceptionV3 with 97.7022% test accuracy. The proposed CNN model had the lowest test accuracy compared to the other models at 94.4852%.

B. Discussion

The reason why the proposed CNN model could be performing better than other transfer learning models could be attributed to insufficient model training. As could be seen from the loss curves of the CNN model, stationarity had yet to be obtained by the end of the 50 epochs. This is different from the transfer learning models, where stationarity is observed towards the end of the 50 epochs. Thus, the proposed CNN model had yet to attain convergence, resulting in lower accuracy. In addition, the transfer learning models used here were created by industry experts in CNN and initialized with pre-trained weights, which have been optimized with the training of the ImageNet dataset. Thus, convergence can be attained earlier with the transfer learning models trained with the new satellite images. The proposed CNN model can be trained with more epochs until convergence is attained, and hyperparameter tuning with more hyperparameters can be performed as part of future work to attain better performance with the model.

5. Conclusion and Future Work

This research introduces a novel approach to road feature classification using convolutional neural networks (CNNs) and transfer learning. For this task, we contrast the accuracy of the suggested CNN model with that of other transfer learning algorithms, including ResNet50, MobileNetV2, VGG19, and InceptionV3. Various hyperparameters, including learning rate and batch size, are investigated in this study. The results demonstrated that both the suggested CNN model and the TL models, which had undergone hyperparameter tuning, can classify roundabouts, crosswalks, intersections, and overpasses with relatively high accuracy. The best-performing model during the validation phase is InceptionV3, with an accuracy of 98.9777%, whereas the best-performing model during the test phase is ResNet50 and VGG-19 models, with an accuracy of 98.7132%. The proposed CNN model got 95.1208% and 94.4852% accuracy during the validation and test stage.

The CNN architecture proved to be well-suited for road feature classification tasks, capturing spatial dependencies and extracting discriminative features from road images. Our model achieved high accuracy, robustness, and efficiency. However, more work needs to be done to improve its classification performance further further.

Our study demonstrates the effectiveness of transfer learning and CNNs in the classification of road features. The proposed approach offers a practical and efficient solution for accurate identification and categorization of road features, contributing to the advancement of intelligent transportation systems and enhancing overall road safety and efficiency.

The implications of our research extend to various domains within intelligent transportation systems, including autonomous driving, traffic management, and road infrastructure maintenance. Accurate classification of road features enables safer navigation, improved traffic flow, and adequate decision-making in transportation planning and management.

Future research can explore additional road feature categories and expand the classification system to handle real-time scenarios. Further investigations can also focus on optimizing the model architecture, refining transfer learning strategies, and incorporating contextual information for more comprehensive road analysis.

References

[1] K. Jurczenia and J. Rak, "A survey of vehicular network systems for road traffic management," *IEEE Access*, vol. 10, pp. 42 365–42 385, 2022.

[2] N. Sakib *et al.*, "Investigating factors influencing pedestrian crosswalk usage behavior in dhaka city using supervised machine learning techniques," *Multimodal Transportation*, vol. 3, no. 1, p. 100108, 2024.

TABLE IV. The Validation Accuracy With diffrent Learning Rate and Batch Size

Trial	Learning Rate	Batch Size	Proposed CNN	InceptionV3	ResNet50	VGG-19	MobileNetV2
1	0.01	32	92.7974	25.2323	25.4647	25.6506	25.9294
2	0.01	64	93.9591	25.4647	32.0167	26.0223	91.5428
3	0.001	32	95.1208	97.4907	95.9572	90.0093	97.3048
4	0.001	64	94.6097	97.9554	96.0967	89.0799	97.2583
5	0.0001	32	91.7286	98.7918	98.5130	97.7695	97.3048
6	0.0001	64	92.1468	98.9777	98.6524	97.9082	98.3271

TABLE V. Validation accuracy and Test Accuracy of Image Classification Models with their Most Optimal Hyperparameters

Model	Hyperparameters	Validation Accuracy (%)	Test Accuracy (%)
Proposed CNN	Learning rate: 0.001, Batch size: 32	95.1208	94.4852
InceptionV3	Learning rate: 0.0001, Batch size: 64	98.9777	97.7022
ResNet50	Learning rate: 0.0001, Batch size: 64	98.6524	98.7132
VGG-19	Learning rate: 0.0001, Batch size: 64	97.9082	98.7132
MobileNetV2	Learning rate: 0.0001, Batch size: 64	98.3271	98.0698

[3] M. Rizzo *et al.*, "Fruit ripeness classification: A survey," *Artificial Intelligence in Agriculture*, vol. 7, pp. 44–57, 2023.

[4] I. H. Sarker, "Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions," *SN Computer Science*, vol. 2, no. 6, p. 420, 2021.

[5] E. Ranyal, A. Sadhu, and K. Jain, "Road condition monitoring using smart sensing and artificial intelligence: A review," *Sensors*, vol. 22, no. 8, p. 3044, 2022.

[6] D. K. Dewangan and S. P. Sahu, "Rcnet: road classification convolutional neural networks for intelligent vehicle system," *Intelligent Service Robotics*, vol. 14, no. 2, pp. 199–214, 2021.

[7] V. Tümen and B. Ergen, "Intersections and crosswalk detection using deep learning and image processing techniques," *Physica A: Statistical Mechanics and its Applications*, vol. 543, p. 123510, 2020.

[8] H. Li, M. Hu, and Y. Huang, "Automatic identification of overpass structures: A method of deep learning," *ISPRS Int J Geoinf*, vol. 8, no. 9, p. 421, 2019.

[9] U. Jilani, M. Asif, M. Rashid, A. A. Siddique, S. M. U. Talha, and M. Aamir, "Traffic congestion classification using gan-based synthetic data augmentation and a novel 5-layer convolutional neural network model," *Electronics (Basel)*, vol. 11, no. 15, p. 2290, 2022.

[10] K. S. Sandeep *et al.*, "Road feature extraction using cnn and gru in edge-enabled vehicular cloud networks," in *2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*. IEEE, 2024.

[11] S. Mekruksavanich *et al.*, "Deep learning approaches for classification of road types using wearable sensors embedded in smart glasses," in *2023 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT NCON)*. IEEE, 2023.

[12] K. Liao, "Road damage intelligent detection with deep learning techniques," in *2022 IEEE 5th International Conference on Information Systems and Computer Aided Education (ICISCAE)*. IEEE, 2022.

[13] K. Bayoudh, F. Hamdaoui, and A. Mtibaa, "Transfer learning based hybrid 2d-3d cnn for traffic sign recognition and semantic road detection applied in advanced driver assistance systems," *Applied Intelligence*, vol. 51, no. 1, pp. 124–142, 2021.

[14] A. A. Alhussan *et al.*, "Pothole and plain road classification using adaptive mutation dipper throated optimization and transfer learning for self driving cars," *IEEE Access*, vol. 10, pp. 84 188–84 211, 2022.

[15] M. Rezapour and K. Ksaibati, "Convolutional neural network for roadside barriers detection: Transfer learning versus non-transfer learning," *Signals*, vol. 2, no. 1, pp. 72–86, 2021.

[16] S. H. Hosseini *et al.*, "Road sign classification using transfer learning and pre-trained cnn models," in *International Conference on Artificial Intelligence and Smart Vehicles*. Springer Nature Switzerland, 2023.

[17] N. Saeed, R. G. Nyberg, and M. Alam, "Gravel road classification based on loose gravel using transfer learning," *International Journal of Pavement Engineering*, vol. 24, no. 2, p. 2138879, 2023.

[18] K. K. M. Shariff *et al.*, "Wet road detection using cnn with transfer learning," in *2022 IEEE 12th Symposium on Computer Applications Industrial Electronics (ISCAIE)*. IEEE, 2022.

[19] X. Qi *et al.*, "Mlrsnet: A multi-label high spatial resolution remote sensing dataset for semantic scene understanding," May 2023, mendeley Data. [Online]. Available: https://data.mendeley.com/datasets/xxxxx/yyyyy

[20] B. Patel, T. D. Barfoot, and A. P. Schoellig, "Visual localization with google earth images for robust global pose estimation of uavs," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 2403–2409.

[21] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Networks*, vol. 106, pp. 249–259, Oct 2018.

[22] D. Dablain, K. N. Jacobson, C. Bellinger, M. Roberts, and N. Chawla, "Understanding cnn fragility when learning with imbalanced data," Oct 2022, available: http://arxiv.org/abs/2210.09465.

[23] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J Big Data*, vol. 6, no. 1, p. 27, Dec 2019.

[24] S. Connor and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of big data*, vol. 6, no. 1, pp. 1–48, 2019.

[25] K. Potdar, T. S. Pardawala, and C. D. Pai, "A comparative study of categorical variable encoding techniques for neural network classifiers," *International journal of computer applications*, vol. 175, no. 4, pp. 7–9, 2017.

[26] M. Mohiuddin, M. S. Islam, and M. H. Kabir, "Performance analysis of bearing fault diagnosis using convolutional neural network," in *2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON)*. IEEE, 2021.

[27] "Recent trends and advances in artificial intelligence and internet of things," 2020, pp. 519–567.

[28] "Image classification using cnn: Introduction and tutorial," https://datagen.tech/guides/image-classification/image-classification-using-cnn/, accessed Jul. 01, 2023.

[29] P. Sharma, "Understanding transfer learning for deep learning," https://www.analyticsvidhya.com/blog/2021/10/understanding-transfer-learning-for-deep-learning/, accessed Jul. 01, 2023.

[30] J. Brownlee, "Understand the impact of learning rate on neural network performance," https://machinelearningmastery.com/understand-the-dynamics-of-learning-rate-on-deep-learning-neural-netw accessed Jul. 01, 2023.

[31] ——, "Difference between a batch and an epoch in a neural network," https://machinelearningmastery.com/difference-between-a-batch-and-an-epoch/#:~:text=The%20batch%20size%20is%20a%20number%20of%20samples%20processed%20before,samples%20in%20the%20training%20dataset., accessed Jul. 01, 2023.

[32] G. Bruno, D. Antonelli, and D. Stadnicka, "Evaluating the effect of learning rate, batch size and assignment strategies on the production performance," *Journal of Industrial and Production Engineering*, vol. 38, no. 2, pp. 137–147, Feb 2021.

[33] I. Kandel and M. Castelli, "The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset," *ICT Express*, vol. 6, no. 4, pp. 312–315, Dec 2020.

Ahmed Abed Mohammed is an Assistant Lecturer at the Department of Technical Engineering, Islamic University. He received his Bachelor's degree in 2020 from Al-Qadisiyah University, Iraq, and his Master's degree in 2022 from University Science Malaysia, Malaysia. His research interests include Machine Learning, Deep Learning, Neural Networks, and NLP. He can be contacted via email at a.alsherbe@gmail.com.

Putra Sumari is a Professor at the School of Computer Sciences, Universiti Sains Malaysia. He received his MSc in 1996 from Liverpool University and PhD in 2001 from John Moores University, England. He leads the robotic, computer vision, and image processing research group at Universiti Sains Malaysia. His research interests include image/video analysis, deep learning, neural networks, and brain-computer interface. He can be contacted via email at putras@usm.my.