# Road State Classification of Bangladesh with Convolutional Neural Network Approach

**Sajid Ahmed**
**Department of Electrical and Computer Engineering, North South University, Bangladesh**
**Dhaka-1229, Bangladesh**

**Taoseef Ishtiak**
**Department of Electrical and Computer Engineering, North South University, Bangladesh**
**Dhaka-1229, Bangladesh**

**Arif Ur Rahaman Chowdhury Suhan**
**Department of Electrical and Computer Engineering, North South University, Bangladesh**
**Dhaka-1229, Bangladesh**

**Mehreen Hossain Anila**
**Department of Electrical and Computer Engineering, North South University, Bangladesh**
**Dhaka-1229, Bangladesh**

**and**

**Tanjila Farah**
**Department of Electrical and Computer Engineering, North South University, Bangladesh**
**Dhaka-1229, Bangladesh**

## ABSTRACT

The Traffic congestion is one of the most intricate and challenging problems in all major cities and urban area of Bangladesh. Inadequate road infrastructure is one of the major causes involved with this agonizing issue. The only existing solution to this issue is manual reporting to authority. This study proposes an app-based road state classification, damage detection, and reporting system to assist both the drivers and authority to identify the damaged roads through a proposed web platform. This paper has made various contributions to address the road type classification of Bangladesh. The proposed research work includes the first of its kind road surface classification dataset, prepared in Bangladesh that could be used for applying machine learning techniques. The dataset has been classified in five classes based on the surface condition. The research team then studied some of the state-of-the-art Residual network based machine learning models and later proposed a customized architecture with a smaller number of layers compared to the state-of-the-art Inception-v3 and Inception-ResNet-V2 architectures for classification purpose. The study has explored three different state-of-the-art machine learning models i.e. Inception-v3, Inception-ResNet-v2, Xception for classification and analyzed their results.

**Keywords**: Inception-v3, Inception-ResNet-v2, Xception, Convolution Neural Network-CNN, Residual block, Fully Connected Network-FCN and Intelligent Transportation System.

## 1. INTRODUCTION

The World Bank has reported that, in the last 10 years, the average traffic speed in Dhaka has dropped from 21 kilometers per hour (kmph) to 7 kmph. This will further reduce by 2035, in 4kmph, which is slower than the walking speed. A study, by BRAC Institute of Government and Development suggests that the traffic congestion in Dhaka wastes up around 5 million working hours every day and costs Bangladesh USD 11.4 billion every year [1]. The cause of this problem lies in various aspects in the country's lifestyle. From inadequate manpower and faulty traffic signaling systems, to overtaking tendency of drivers and narrow roads, from heavy concentration of vehicles, absence of adequate public transport, to inadequate road infrastructure and poor enforcement of traffic rules, moreover over population in city areas and people disrespect towards traffic rules add up to this traffic chaos in the cities of Bangladesh. The government with various other organizations such as UN, UNDP, and BRAC has undertaken various projects to tackle this traffic congestion. These projects include expansion of footpaths, and building flyovers, overpasses ring roads to deviate traffic from the city center, metro rail lines, rapid bus routes, new roadways and new traffic rules for drivers and pedestrians [2]. While all these initiatives are commendable, making them sustainable should also be taken into consideration. The biggest challenge Bangladesh faces is the over population.

Any infrastructure government build would have to handle more people traffic then most of the countries in the world. Thus, maintaining good infrastructure condition especially roads is a challenging duty. In most cases while government completes building one infrastructure another one falls apart due to lack of maintenance and the traffic congestion increases. Various socio-economic research has been conducted to address the issues relating to traffic congestion and how to handle them. But there hasn't been much noticeable research done on technical aspects and modern tools that could be used to address the traffic congestion and infrastructure sustainability issues in Bangladesh.

This study aims to use image processing and classification through machine learning to identify road surface condition in Bangladesh. These classifications are then used in a web application along with google map API for GPS location information to notify the users both the location of the road and road surface condition. Thus, if a certain area's road is damaged or under construction drivers or newcomers could decide whether to take the road or how to travel in that road. This web application depends on crowd sourced data for new road images. Once a road image is classified among one of the classes of Perfect Road, Good Road, Mild Bad Road, Severely Bad Road or Water on Surface Road condition, it will notify the end user about the road surface condition accordingly.

There have been several studies on classification of road state image using various machine learning models. But most of those researches used road image of developed countries. The average condition of roads vary country wise. In developing and underdeveloped countries; like Bangladesh, different roads are formed with different building materials. For the building material of roads, the surface condition and change in structural health of roads vary. For the classification purpose a dataset of more than 5000 road surface images of various roads of Bangladesh had been prepared in this study. Analyzing 2D RGB images of the road surface condition of Bangladesh from the dataset and various machine learning models were trained to find the pattern to classify the roads based on surface condition. The most widely used machine learning model for image classification is convolutional neural network (CNN). Classical deep Neural Networks like CNN and FCN extract image features finding key points and thus used for 2D image classification. But Google's Inception-v3 architecture has argued to reduce the computational cost using residual networks. This study has explored several models built with deep residual blocks for road image classification and makes the performance comparison among the models [3]. This model is modified to classify the 2D road images and ensure computational efficiency in lighter resources. The study also shows a comparison among Inception-v3, Inception-ResNet-v2, Xception and justify the choice to build a light customized CNN architecture, inspired from the architectures of Inception-v3 and Inception-ResNet-v2 for classification purpose. Smart ITS systems are needed for countries like Bangladesh. The proposed model architecture can be implemented for autonomous car technologies for developing countries, where the vehicle can get idea to monitor its speed and direction analyzing the surface condition; and, in advanced robotics industries.

This paper is organized as following: Section 2 discusses about all the related work done is this field of research. In Section 3 the methodology of used in conducting this research has been discussed. In Section 4, the machine learning models used in this study has been explored. The dataset collection and screening labelling and classification process of this study has been explained in Section 5. The results of this study have been discussed in Section 6. The web and mobile platform of this research has been discussed in Section 7. The paper is concluded in Section 8.

## 2. LITERATURE REVIEW

There have been various studies on the road-traffic conditions, causes and their affects in the socio-economic, physical and mental health the people of Bangladesh [1][2][4]. None of these studies suggested any technical solution that this paper is proposing. CNN based machine learning models are being very effectively used for pattern recognition, image classification, object detection, semantic and instance segmentation. So far, the "AlexNet" architecture introduced by Krizhevsky Et al. [5] has been successfully applied in a number of computer vision tasks, for example in object detection [6], object tracking [7], segmentation, [8] video classification, [9] human pose estimation [10] and super-resolution [11]. Classical deep Neural Networks like CNN and FCN extract image features finding key points and thus are used for 2D image classification. On the other hand, Google introduced their own residual deep neural network architecture in Inception-v1, they have optimized the network furthermore through Inception-v3 to reduce computational cost [10][11]. The basic difference between Inception-v3 and Inception-v4 is that the latter one has a simpler architecture and more inception modules [12]. The constraints contained in Inception-v3 had come from the need for partitioning model architecture with distributed training using DistBelief [27]. With migration to TensorFlow [13] those constraints have been lifted which allowed the reintroduced architecture to get simplified. The goal of this research was to analyze the road surface condition of Bangladesh from 2D images and train a model to find the pattern to classify the roads based on surface. Based on the collected dataset preprocessing with canny edge detection, image black and white thresholding, Fourier transformation, the road images were classified into 5 classes: Perfect Roads, Good Roads, Mild Bad Roads, Severely Bad Roads, and Water on Surface Roads. For this classification, a convolutional neural network approach and a combined modification of Google's Inception-v3 and Inception-ResNet-v2 model architectures were used to classify 2D images of roads have been used. CNN has been successfully used in 2D image classification and image-processing based research such as: Food Classification [14] and Gaussian noise detection [15]. For this reason, CNN was used to classify road state images. The trained model in this study analyzed the change of pixel intensity for the existence of cracks, garbage on roadsides, pits, surface water, and road-color change for hilly tracts and soil change. The acquired dataset images were taken on almost keeping the same frame ratio of images. This approach helped the machine learn about the width of the road. At the same time the dataset was labeled accordingly. CNN based researches have been performed to detect road, semantic segmentation of road scenes [16-20], road-lane, road area extraction [21], rural roads detection [22], street signs detection and so on. The works are most commonly based on classification, object detection or semantic segmentation. For image recognition and object-detection based works some of the novel approaches have been introduced by Lin et al. [23], Simonyan et al. in VGGNet [24] and GoogleNet (Inception-v1) [25] by Szededy Et al. Residual connections were introduced in [3] where the authors emphasized for the practical advantages of using additive merging of signals in image processing and object detection. Computational power efficiency [26] is a very good approach for real-time joint de-noising of images, vehicle detection and road segmentation. For surface health monitoring, in civil engineering projects, deep convolutional neural network approaches like; bounding box approaches have been proposed in computer vision-based models. A pothole detection model likewise; had been proposed in [27] based on semantic segmentation on concrete crack images which is also a fully connected network (FCN) approach. Road object recognition with semantic segmentation have been one of the most advanced field of research. Several semantic image segmentation models have been introduced there by [28-31].

## 3. METHODOLOGY

As a developing country, the roads in Bangladesh varies in structure and construction materials. Thus, the existing road image dataset and trained models for road condition analysis is inefficient for classifying the road images collected in Bangladesh. There are several road state analysis systems for developed countries. But they don't have sufficient data for classifying developing or underdeveloped countries road surface condition. For this study the dataset used in previous research has merged with the collected dataset of Bangladeshi roads. For data preprocessing, the data was reshaped to images of 128 by 128 size. Several image analysis techniques: canny edge implementation, black and white thresholding and Fourier Transformation of image signals were executed for splitting the images into 5 training classes. The training images were of similar frame size and taken during bright daylight. For finding the best classifier model for road surface image classification, the dataset was trained, validated and tested with several existing state of art machine learning models.

The state-of-the-art models have been very efficient for classifying dataset with large number of classes. But in this study only five classes were selected for classification. Among the 3 state of art models, Inception-v3 and Inception-ResNet-v2 showed observable satisfactory results. Inspired by the results of these two models a customized model was implemented for obtaining classification and which has been proved computationally efficient for the proposed research work. The basic classification of training set was performed, and the images were labeled based on the parameters mentioned in Table 1. The training images were labelled to analyze the cracks and other anomalies present on road surface observing the RGB image and using OpenCV libraries. After image preprocessing, of the training set images, the images have been analyzed for variance over surface pixel values achieved through Fourier transformation. The imperfect roads that we considered from our dataset based on our image analysis mentioned above; are unfit for human transportation. Following the features of Table 1, the research team has analyzed the density of cracks, width of roads, surface smoothness, color of roads, environmental metric present on roads such as: presence of water due to rainfall, sand or raw soil on road surface and trained the model on such scenario.

Table 1 Parameters Considered to Detect the Road Condition

| Sl. No. | Parameters |
|---|---|
| 1. | Cracks severity |
| 2. | Road width |
| 3. | Surface color |
| 4. | Weather effect on road |
| 5. | Vibration felt due to cracks |
| 6. | Drainage, garbage and other anomalies presence |

The customized model is used in the backend of a web application for running a web platform where a user can upload a particular road's image. During the picture snap taking with the fixed frame size with the help of Android Mobile Application described in section 7, the GPS coordinates are taken. This allows to prevent false data to entry in the database. The image gets automatically classified using the proposed customized model's weight file in the backend. The other end user when searches the location, gets the image and road surface condition through this web platform. This web application thus can be used by user for checking current road condition and uploading new road image for future references. Every time a new image is uploaded in the web platform, the model classifies the image and the new images can be used for future training and validation.

## 4. MACHINE LEARNING MODELS

To ensure performance with respect to the corresponding dataset, several models were trained and tested. Among them only three different state-of-the-art models gave good performance and hence described in this section. Based on this experience, a customized CNN model has been built. The common element in all models is residual block.

Residual block was first introduced in [3] for building very deep neural network architectures for image recognition which later have been developed to have deep layer architectures of up to 152 layers. Introducing the skip connection paradigm residual network has enabled to have deeper network and solve the vanishing gradient problem. The layers in residual network thus learn the residual representation functions instead of learning the signal representation directly. Kaiming He Et al. solved the counter intuitive problem faced by classical deep learning networks that when the network becomes large in depth the problem of vanishing gradient occurs [11]. During backpropagation, the weights of a model in each iteration of training are used for partial derivative calculation of the error function with respect to the current weight. If a large n-layer network is deep enough to be multiplied for n-times with small partial derivative values to compute gradients, the gradients eventually become zero. Oppositely, if the n-layer network is deep enough and the partial derivative values are large then the multiplication o n-times make the gradients too large, eventually exploded. The concept of skip connection solves this issue and enables to build very deep layers.

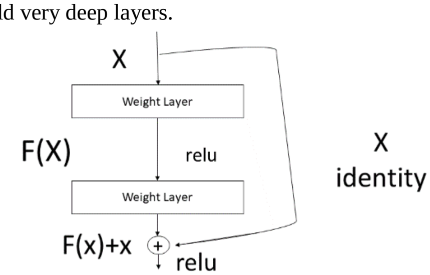

Figure 1: Building block of residual network

In skip connections, the input x is added to the output after few layers like the pictorial representation presented in Figure 1. Therefore, the actual output of a building block $H(x) = F(X) + x$ learns a kind of residual mapping with $F(x) = H(x) – x$ as represented in Figure 1. When the vanishing gradient occurs, the identity x transfers back to the earlier layer information. To remove high time complexity related bottleneck, the 1 x 1 conv layers have been suggested to add in [23][25]. This 1 x 1 conv later can reduce the number of parameters without degrading the performance of the model.

Among the variants of residual network based models, the corresponding dataset of this research has been trained on several variants of ResNet which are: Xception, Inception-ResNet-v2 and Inception-v3. The performance comparison has been presented in section 6 of this paper. This paper therefore, will focus on the performance comparison analysis upon the dataset in terms of training efficiency.

**Xception**
Another depth wise separable convolution had been introduced in the extreme version of Inception network, which is commonly known as the Xception architecture [32]. The Xception network is proven to be better than Inception-v3 tested upon the ILSVRC and JFT datasets. The original depth wise convolution is traditionally followed by a pointwise convolution. The Inception-v3 architecture has 1x1 convolution performed before n x n spatial convolution. Whereas the original depth wise separable convolutions perform the channel-wise spatial convolution first and then perform the 1x1 convolution. The modified depth wise separable convolution does the opposite, but this approach does not make any significant difference in performance between Inception-v3 and Xception. Also, another difference from Inception-v3 is the absence of non-linearity intermediate activation of ReLU.

### Inception-Resnet-v2

The Inception-Resnet architecture proposed in [33] was initially motivated to find if there is any benefit in combining Inception architecture with residual architecture. Later they proved with two separate variants naming Inception-ResNet-v1 and Inception-ResNet-v2 that, the residual block significantly accelerates the training of inception architecture and found residual-inception networks outperforming inception networks without residual connection by a thin margin. It is stated that several variants of Inception-ResNet architectures have been tried in the work but two variants Inception-ResNet-v1 and Inception-ResNet-v2 have been explained in the paper. The Inception-ResNet-v1 roughly meets the computational cost on Inception-v3 while the Inception-ResNet-v2 matches the computational cost of Inception-v4.

### Inception-v3

The inception architecture was rethought in Inception-v3 with fewer parameters for computational efficiency [12]. It was found that with 42-layer deep learning architecture, similar complexity like VGGNet was achieved. The third version of the Inception model, the idea of factorization was brought, which main objective was to reduce the number of parameters without losing network efficiency. With 1 layer of 5x5 filter size and 2 layers of 3x3 filters, the number of parameters was reduced by 28%. With this technique a new Inception module was introduced like the representation shown in Figure 2.

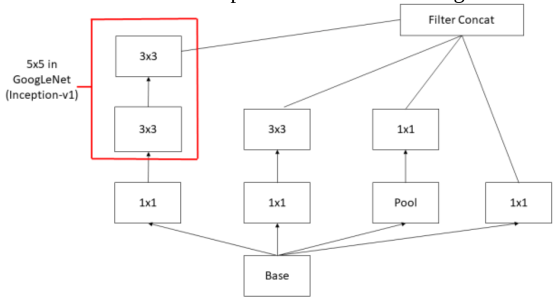


Figure 2: Inception module using factorization

Later the 3x3 convolution was replaced by single 3x1 convolution followed and 1x3 convolution. This approach reduced the number of parameters by 33%. This technique brought one of the new Inception modules represented in the Figure 3.

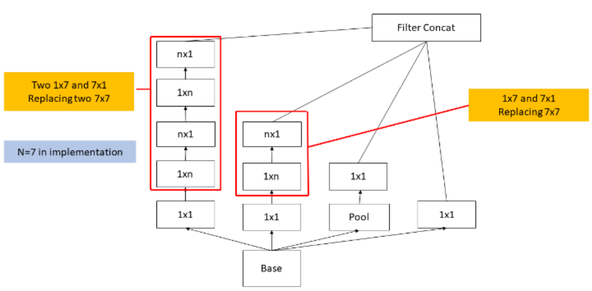


Figure 3: Inception module using asymmetric factorization

For high dimensional representations described in [11], another Inception module was introduced, which is represented in Figure 4. These three Inception modules with factorization was useful to reduce the number of parameters for the whole network and it was able to avoid overfitting and the network can be of deeper architecture. As suggested in [26] Auxiliary classifier had been used in Inception-v3, but instead of two units, single unit on the top of the last 17x17 layer. This auxiliary classifier was used as regularizer. An efficient grid size reduction was implemented to avoid the greedy approach of max pooling followed by conv later and the expensive approach of the conv layer followed by max pooling.

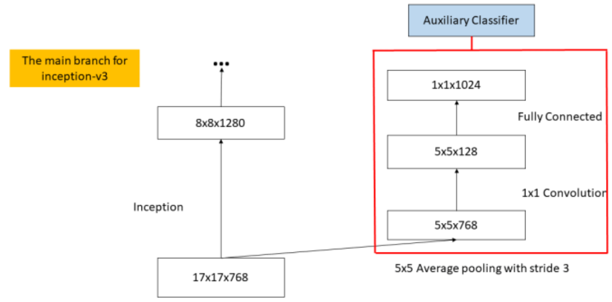


Figure 4: Auxiliary Classifier

### Customized CNN Model

Among the explored state-of-the-art models, the Inception-v3 and Inception-ResNet-v2 models gave the best performance on the acquired dataset. For further analysis the research team continued to make a custom model which outperformed all other models. The customized model architecture was encouraged from the Inception-v3 and Inception-ResNet-v2 architectures. The customized model architecture consists of 32 by 32 convolutional layers containing 5 output layers. Later 32 size filter, 2D MaxPooling with pool size of 2 by 2 had been used. This part was taken from the Inception-v3 [12] architecture. For batch normalization, ReLU and Softmax activation was used inspired from Inception-ResNet-v2 [33].

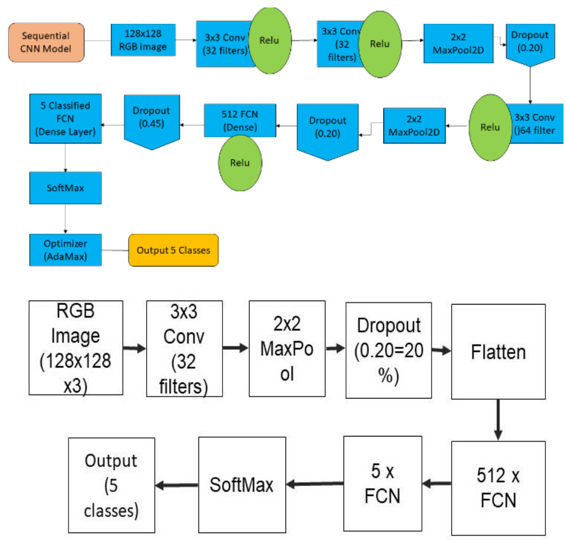


Figure 5. Customized CNN Model Architecture.

In the proposed residual model architecture, the convolution block consists of total 13 layers and has been divided into 3 residual blocks. The input images' size was converted to 128 by 128 pixels. In the first block it has two 3 by 3 convolution layers with 32 filters and one 2 by 2 MaxPooling layer and a dropout of 20 percent. On the second block the model has one 3 by 3 Convolution layer with 64 filters and one 2 by 2 MaxPooling layer and a dropout of 20 percent. The third block has three fully connected layers with dense 512 neurons and a dropout of 45 percent. Finally, it has a classified layer to classify the images into 5 different classes. After each and every convolutional block 'ReLU' was used as the activation function. SoftMax activation was used at the final classification layer. To get an optimized output 'Adamax' was used as optimizer with 'categorical_crossentropy' loss function to calculate loss. The configuration of our training resource for the residual architecture is same as mentioned above for CNN based FCN architecture. Classical CNN gets trained with image feature learning. With steps by steps of convolution layer and with proper activation function, the network extracts the key features from image pixel boxes and starts to recognize the pattern.

But the old Inception model was trained in a partitioning manner. Each replica of that network was partitioned into sub-networks for the memory to fit the whole network. With the introduction of TensorFlow [13], the most recent Inception models can be trained without the replica partitioning. Recent optimizations of memory with backpropagation enabled in part and it achieved considering what tensors are needed for gradient computation and trying to reduce such tensors by structuring the computation. As our research team was experimenting on how a residual deep network works comparing to another classical CNN architecture and achieved better results, described in section 6, the Figure 5 demonstrates our modified ResNet and CNN architecture that we ended up considering based on our achieved best result.

The configuration of our training resource is Intel core i5 8th gen 6 core @4.00 GHz 32 GB DDR4 RAM PC NVidia GTX 1070 8 GB GDDR5 Memory: Dedicated 8GB + 8 GB Shared 16 GB GPU Memory in total. The training time for 50 epochs was 30.39 minutes for training 4,430 images. To analyze the images perfectly for splitting of training set, black and white thresholding was applied before splitting to confirm the existence of anomaly on road surface and cracks, moreover, canny edge detection, Fourier analysis were also applied. This process was human observation based and images were decided to split into 5 classes from the observation. The research team considered the following parameters present in Table 1 to analyze our images.

The training set data at last had been fed into the proposed Inception-v3 and Inception-ResNet-v2 based deep neural network with the FCN network. The network learns how the surface features from Table 1 are present on an image and learns the color pattern of the road types.

## 5. DATA PROCESSING

Data collection, screening, labelling and classification were the most diligent work of this project. As mentioned earlier this is the first dataset for road surface condition of Bangladesh. The images were captured from above the road or from an onboard vehicle camera as shown in Figure 6. Thus, the trained model can classify images taken from any angle. This will be helpful when the model will classify the data uploaded from user. Our dataset consists of 5,370 images. Among them 4,430 images have been kept for training data split into 5 classes as mentioned in Table 2 with the split mentioned in Table 3; 340 of the other remaining images have been kept for validation set and rest 600 images have been kept for test set. Among the whole dataset, 3,785 images were collected by the research team over a time period of 4 months, the remaining 1,285 images have been acquired from Local Government Engineering Department (LGED) of Government of People's Republic of Bangladesh [34] and 300 best quality images were collected from google images [35].

For the classification of validation and testing set we considered all the factors mentioned in Table 1. Figure 7 gives a brief idea about what types of images the dataset holds. The statistics of train, validation and test image set is compiled on Table 2.

### B) Training Classes Description:

**1) Perfect Roads:** The best category concrete RCC and Rigid roads have been taken for training under this training class. Roads that have a smooth surface condition and no cracks found have been taken for consideration. These roads are very wide and mostly of highway roads of the country.

Figure 6: Data collection with Smartphone installed on a car.

Table 2. Dataset description overview

| Type | Class Name | Detail |
|---|---|---|
| Crack | Perfect Road | smooth surface condition and no cracks (wide and highway roads) |
| Crack | Good road | roads are narrower than RCC and Rigid roads |
| Crack | Mild Bad Road | crack parts are visible in the images |
| Crack | Severely Bad Road | big cracks, pits muddy condition |
| Conditi on | Water on Surface | visual water on the whole road surface |

Table 3. Dataset Split.

| Classes | Training | Validation | Test |
|---|---|---|---|
| Perfect Road | 865 | 68 | 150 |
| Good Road | 925 | 68 | 150 |
| Mild Bad Road | 1,360 | 68 | 130 |
| Severely Bad Road | 700 | 68 | 100 |
| Water on Surface | 580 | 68 | 70 |

**2) Good Roads:** The local area roads which are good in condition but structurally not sustainable as RCC and Rigid roads are have been taken for training in this category. These roads are narrower than RCC and Rigid roads. Some roads of the hilly tracts of the country which are good for transportation have been taken for consideration in this class as well.

**3) Mild Bad Roads:** The road images which surface have been decayed but not that much bad in condition have been taken for training under this class. The crack parts are visible in the images and which have been confirmed with applying OpenCV's black and white thresholding as shown in Figure 7.

**4) Severely Bad Road:** The road surface condition in this class has been considered unfit for transportation. The concrete carpet has been totally removed by environmental and heavy transportation reasons. There are big cracks, pits muddy condition visible in the images for such condition.

**5) Water on Surface Road:** Roads that have standstill water on surface have been considered for training. The images taken for this training class have visual water on the whole road surface, but not on the pits only. Similar work has been done in [36] but for our proposed research work roads that have whole area drowned beneath water as shown in Figure 7.


Figure 7: Dataset abstract view: Sample 5 classes of Roads

For verifying the performance of customized CNN proposed in this study, it was tested with few of the famous dataset i.e. Cityscapes and KITTI road [37][38].

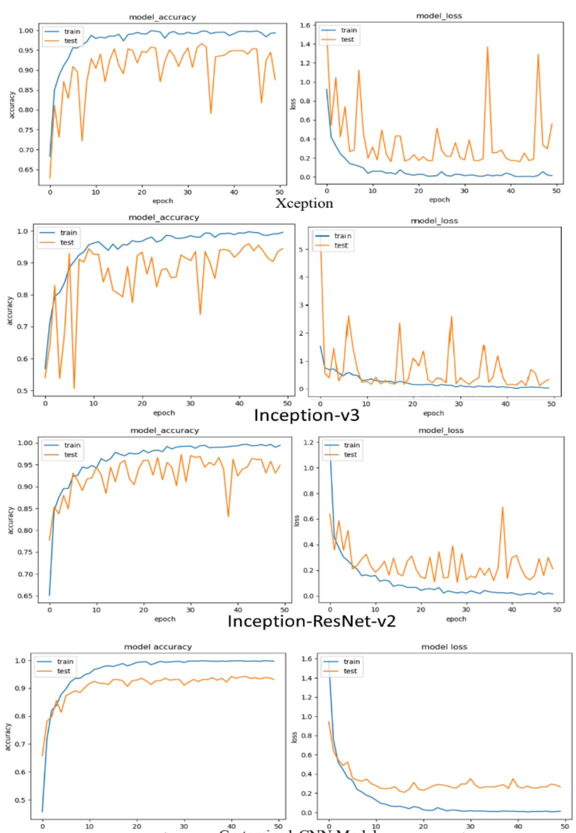


Figure 8: Graphs of the explored models' accuracy and loss showing respectively: Xception, Inception-v3, Inception-ResNet-v2, and proposed customized CNN model.

## 6. RESULT ANALYSIS

To validate the best model for this system the dataset has tested with various models. Both the collected dataset and the Cityscapes [37] and KITTI [38] road dataset have been used for model validation. Each model was trained for 50 epochs and the test accuracy graph differs model wise. The performance against training and test-validation set of the explored models can be interpreted from the Figure 8.

The Xception model did not show satisfactory performance against the test set whereas, the Inception-ResNet-v2 and Inception-v3 showed better results which can be observed from Figure 8. From this intuition, the customized CNN model was implemented based on the architecture of Inception-ResNet-v2 and Inception-v3. However, based on the test accuracy described in Table 4 the customized CNN model gave the best result as shown in Figure 8. The model was experimented for 0 to 50 epochs and the accuracy sharply raised till 8th epoch and gradually flattened till 50 epochs which is also observable in Figure 8.

For computational efficiency, the customized architecture contains 8 layers instead of 48 layers of Inception-v3 or 164 layers of Inception-Resnet-v2. The customized model took around 30 minutes for training where the dense layered original state of art architectures took more than two hours each. The result of classification on the local dataset acquired by the research team, is mentioned on the road images as shown in Figure 9 and Figure 10. These images are system generated results for customized CNN model as this model has been used in web and mobile platform based on the test accuracy result. For verification, if the model proposed in this study works for all dataset, some pictures of Cityscapes and KITTI road dataset have been tested with this model and the results are shown in Figure 10. As Cityscapes and KITTI road dataset do not have images of all the classes categorized in this study, only two classes of images have been classified and the corresponding result images are shown in Figure 10.


Figure 9: Test Set results of classification on acquired local dataset.

Table 4: The explored models' Test results

| Model Name | Road Type | TN | FP | FN | TP | Precision | Recall | Accuracy | F1 Score |
|---|---|---|---|---|---|---|---|---|---|
| Customized CNN | Severely Bad Road | 45 | 2 | 2 | 7 | 0.777778 | 0.777778 | 0.928571 | 0.777778 |
| | Perfect Road | 33 | 3 | 2 | 18 | 0.857143 | 0.9 | 0.910714 | 0.878049 |
| | Mild Bad Road | 45 | 1 | 2 | 8 | 0.888889 | 0.8 | 0.946429 | 0.842105 |
| | Mild Good Road | 43 | 2 | 3 | 8 | 0.8 | 0.727273 | 0.910714 | 0.761905 |
| | Water on Surface | 48 | 2 | 1 | 5 | 0.714286 | 0.833333 | 0.946429 | 0.769231 |
| Inception-ResNet-v2 | Severely Bad Road | 47 | 0 | 6 | 3 | 1 | 0.333333 | 0.892857 | 0.5 |
| | Perfect Road | 29 | 7 | 3 | 17 | 0.708333 | 0.85 | 0.821429 | 0.772727 |
| | Mild Bad Road | 42 | 4 | 2 | 8 | 0.666667 | 0.8 | 0.892857 | 0.727273 |
| | Mild Good Road | 40 | 5 | 6 | 5 | 0.5 | 0.454545 | 0.803571 | 0.47619 |
| | Water on Surface | 49 | 1 | 0 | 6 | 0.857143 | 1 | 0.982143 | 0.923077 |
| Inception-v3 | Severely Bad Road | 45 | 2 | 8 | 1 | 0.333333 | 0.111111 | 0.821429 | 0.166667 |
| | Perfect Road | 30 | 6 | 4 | 16 | 0.727273 | 0.8 | 0.821429 | 0.761905 |
| | Mild Bad Road | 36 | 10 | 2 | 8 | 0.444444 | 0.8 | 0.785714 | 0.571429 |
| | Mild Good Road | 40 | 5 | 8 | 3 | 0.375 | 0.272727 | 0.767857 | 0.315789 |
| | Water on Surface | 50 | 0 | 1 | 5 | 1 | 0.833333 | 0.982143 | 0.909091 |
| Xception | Severely Bad Road | 45 | 2 | 6 | 3 | 0.6 | 0.333333 | 0.857143 | 0.428571 |
| | Perfect Road | 31 | 5 | 5 | 15 | 0.75 | 0.75 | 0.821429 | 0.75 |
| | Mild Bad Road | 41 | 5 | 3 | 7 | 0.583333 | 0.7 | 0.857143 | 0.636364 |
| | Mild Good Road | 36 | 9 | 6 | 5 | 0.357143 | 0.454545 | 0.732143 | 0.4 |
| | Water on Surface | 50 | 0 | 1 | 5 | 1 | 0.833333 | 0.982143 | 0.909091 |


Figure 10: Test Set results of classification based on the Cityscapes and KITTI road dataset.

The explored models' testing performances are compiled in Table 4 with the representation of precision and recall score. The average accuracy for customized CNN model outperformed other models. Among the mentioned class types in Table 1, due to lack of available dataset the test classes were not evenly distributed with equal number of images. Table 4 has been generated with a sample portion of the test set images. Only the customized CNN model architecture showed consistent results along the 5 classes with consistent values. Although the precision-recall score came lower for the "Water on Surface" road condition for proposed customized CNN model, while the other state-of-the-art models performed better for "Water on Surface class". But with respect to other classes the overall performance for other models are not satisfactory. Moreover, the precision-recall score across the classes fluctuate verily for other models except for the customized CNN architecture.

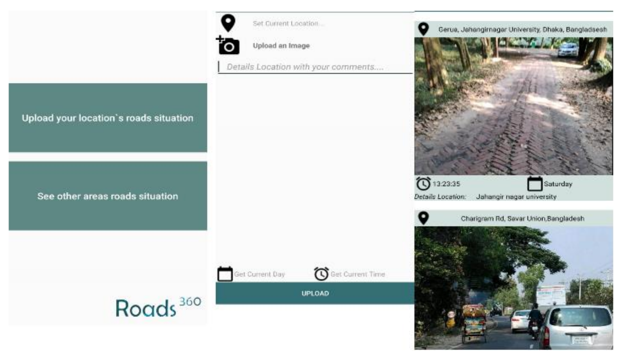


Figure 11: Mobile application interfaces

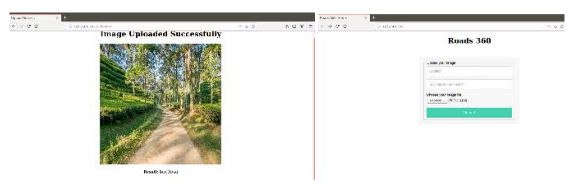


Figure 12: Web application interfaces

## 7. WEB PLATFORM

Initial research plan included making this system available for everyone for free. In next step this platform would be proposed with google map as an add-on so that the user can see the road condition whenever they search for a location on Google Maps. The user interfaces of mobile application and web platform are shown in Figure 11 and Figure 12. Such platforms can be used for uploading condition images to share the surface condition. The customized CNN model weight has been implemented in the backend of a web application where the uploaded image from the android application gets automatically tested without any human intervention and gets classified updating in the database accordingly. This system helps the end user to get the images of a particular place's road condition. Whenever a user uploads an image using any of these platforms, the current location of the user is also recorded. The backend model can analyzes the image and identify the road condition. According to the classified label the images gets automatically updated in the database corresponding the GPS location. Later when any other user makes a search for that location image, the mobile application shows the latest image with road condition mentioned. The platform can also be used by the Government authority of any country as the platform is collecting and sorting various places' road pictures accordingly. As a result, the authority can get the information about the places where the roads surface is in bad condition and needs to be fixed urgently.

## 8. CONCLUSIONS

The motivation of this research work is to build an efficient machine learning image classification model that have a decent accuracy to classify the roads in Bangladesh and alike roads of other developing and developed countries. Existing road state classifiers could only identify the roads of developed country. Dataset for road state of developing country was also unavailable. In this study first ever road surface state dataset for roads of Bangladesh have been prepared. The dataset includes images of ideal good concrete Rigid and RCC roads and also rural roads of Bangladesh making the dataset very versatile. For making the system importable for users a customized CNN model was implemented which takes less time for training and gives more accurate results than the current state of art models. A web & mobile interface was also developed to open the model for public use, so that more road image data can be achieved by our research team to check the model performance across varieties of image framing and daylight-based conditions. The customized model was tested by few other benchmark datasets and provided satisfactory results. The overall system can be proposed to be used as an add-on with Google Maps to achieve additional information about road surface of any particular location along with the location's traffic, navigation and other data.

## 8. REFERENCES


[1] A. A. Haider, "Traffic jam: The ugly side of Dhaka's development," [Online], Available: https://www.thedaily star .net /opinion/society/traffic-jam-the-ugly-side-dhakas-development-1575355. [Accessed Dec 14, 2019].

[2] K. Mahmud, K. Gope, S. Mustafizur, and S. Chowdhury, "Possible Causes & Solutions of Traffic Jam and Their Impact on the Economy of Dhaka City" **Journal of Management and Sustainability**, Volume 2, 2012.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", **arXiv preprint arXiv:1512.03385,** 2015.

[4] R. Shamsher, and M. Abdullah, "Traffic Congestion in Bangladesh- Causes and Solutions: A study of Chittagong Metropolitan City," **Asian Business Review**, Volume 2, March 2015.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks", **In Advances in neural information processing systems, pages** 1097–1105, 2012

[6] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation", In **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),** 2014.

[7] N. Wang and D.-Y. Yeung, "Learning a deep compact image representation for visual tracking", In ***Advances in Neural Information Processing Systems,*** pages 809–817, 2013.

[8] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation", In **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,** pages 3431–3440, 2015.

[9] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks", In **Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, pages 1725–1732. IEEE,** 2014.

[10] A. Toshev and C. Szegedy, "Deeppose: Human pose estimation via deep neural networks", In **Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference** on, pages 1653–1660. IEEE, 2014.

[11] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution", In **Computer Vision–ECCV 2014, pages 184–199. Springer,** 2014.

[12] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision", **arXiv preprint arXiv:1512.00567,** 2015.

[13] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, ´ R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. War- ´den, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. **TensorFlow: Large-scale machine learning on heterogeneous systems, 2015.** Software available from tensorflow.org.

[14] A. Goswami and H. Liu, "Deep Dish: Deep Learning for Classifying Food Dishes,"(2017) [Online], Available: **https://www.semanticscholar.org/paper/Deep-Dish-%3A-DeepLearning-for-Classifying-Food-Goswami Microsoft/2e596abd9d4616f75c9da45d88787cb2fd5bee81.** [Accessed April 29, 2019].

[15] J. H. Chuah, H. Y. Khaw, F. C. Soon and C. Chow, "Detection of Gaussian Noise and Its Level using Deep Convolutional Neural Network", **Proc. of the 2017 IEEE Region 10 Conference (TENCON), Malaysia,** November 5-8, 2017.

[16] M. Teichmann, M. Weber, M. Zöllner, R. Cipolla and R. Urtasun, "MultiNet: Real-time Joint Semantic Reasoning for Autonomous Driving", **IEEE Intelligent Vehicles Symposium, IV 2018, Changshu, China,** June 26-30, 2018, pp. 1013-1020.

[17] Y. Lyu, and X. Huang, "Road Segmentation Using CNN with GRU", **22nd International Conference on Information Visualization,** 2018.

[18] Z. Zhang, Q. Liu, Y. Wang. "Road extraction by deep residual U-net", **IEEE Geosci. Remote Sens. Lett., vol. 15, no. 5, pp. 749-753,** May 2018.

[19] J. G. Haran, J. Dillenburg, and P. Nelson, "Realtime Image Processing Algorithms for the Detection of Road and Environmental Conditions", **Ninth International Conference on Applications of Advanced Technology in Transportation, AATT 2006, Chicago, Illinois,** August 13-15, 2006, pp. 55-60.

[20] C. V. Dung and L. D. Anh, "Autonomous concrete crack detection using deep fully convolutional neural network", **Automation in Construction,** Volume 99, Pages 52-58, March 2019.

[21] J. Zhao, H. Wu, and L. Chen, “Road Surface State Recognition Based on SVM Optimization and Image Segmentation Processing,” **Journal of Advanced Transportation,** Vol 2017, pp. 1-21. 2017.

[22] Z. Wan-zhi, W. Zeng-cai, “Rural Road Detection of Color Image in Complicated Environment”, **International Journal of Signal Processing, Image Processing and Pattern Recognition**, Vol. 6, No. 6 (2013), p.p. 161-168.

[23] M. Lin, Q. Chen, and S. Yan, “Network in network”, **arXiv preprint arXiv:1312.4400,** 2013.

[24] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition”, **arXiv preprint arXiv:1409.1556, 2014**.

[25] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions”, In **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,** pages 1–9, 2015.

[26] J. Dean, G. Corrado, R. Monga, K. Chen, M. Devin, M. Mao, A. Senior, P. Tucker, K. Yang, Q. V. Le, et al. “Large scale distributed deep networks”, In **Advances in Neural Information Processing Systems,** pages 1223–1231, 2012.

[27] J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden, and H. Balakrishnan, “The pothole patrol: using a mobile sensor network for road surface monitoring,” In **Proceedings of the 6th international conference on Mobile systems, applications, and services (MobiSys '08), New York, USA,** June 17 - 20, pp. 29-39.

[28] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, “Pyramid Scene Parsing Network”, **arXiv preprint arXiv:1612.01105v2 [cs.CV] 27 Apr 2017.**

[29] V. Badrinarayanan, A. Kendall, & R. Cipolla (2017), Segnet: A deep convolutional encoder-decoder architecture for image segmentation, **IEEE transactions on pattern analysis and machine intelligence, 39(12), 2481-2495**.

[30] M. Teichmann, M. Weber, M. Zoellner, R. Cipolla, & R Urtasun, (2018, June), Multinet: Real-time joint semantic reasoning for autonomous driving. In 2018 **IEEE Intelligent Vehicles Symposium** (IV) (pp. 1013-1020). IEEE.

[31] T. Pohlen, A. Hermans, M. Mathias, B. Leibe, “Full-Resolution Residual Networks for Semantic Segmentation in Street Scenes”, **arXiv:1611.08323v2 [cs.CV] 6 Dec 2016.**

[32] F. Chollet, “Xception: Deep Learning with Depthwise Separable Convolutions”, **arXiv preprint arXiv:1610.02357v3 [cs.CV]** 4 Apr 2017

[33] C. Szegedy., S. Ioffe, Vincent V., A. Vanhoucke. “Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning”, **arXiv preprint arXiv:1602.07261v2 [cs.CV]** 23 Aug 2016

[34] Local Government Engineering Department (LGED), **http://www.lged.gov.bd/** [Accessed November 11, 2018].

[35] Google Images, **http://www.images.google.com** [Accessed February 25, 2019].

[36] K. Ueda, I. Horiba, and K. ledaga, “A detecting method of wet condition on road using image processing”, **Information Processing Society of Japan,** vol. 35, no. 6, pp. 1072–1080, 1994.

[37] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, Schiele, B. (2016), The cityscapes dataset for semantic urban scene understanding, **In Proceedings of the IEEE conference on computer vision and pattern recognition**, (pp. 3213-3223).

[38] J. Fritsch, T. Kuehnl, & A. Geiger (2013, October), A new performance measure and evaluation benchmark for road detection algorithms, **In 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)**, (pp. 1693-1700). IEEE.