

# Off-road terrain classification

Lafras Fritz<sup>a</sup>, Herman A. Hamersma<sup>b\*</sup>, Theunis R. Botha<sup>c</sup>

Department of Mechanical and Aeronautical Engineering, University of Pretoria, South Africa

<sup>a</sup>[fritzlafras@gmail.com](mailto:fritzlafras@gmail.com), <sup>b</sup>[hermanh@up.ac.za](mailto:hermanh@up.ac.za), <sup>c</sup>[theunis.r.botha@protonmail.com](mailto:theunis.r.botha@protonmail.com)

## ABSTRACT

Road traffic accidents place a burden on the global economy. This impact is reduced by the development of safer vehicles. Advanced Driver Assist Systems (ADAS) aim to reduce the frequency and severity of accidents. ADASs are designed to operate in well-defined environments, such as first world urban areas. However, 93% of fatal accidents occur in developing countries; areas often without properly maintained roads. ADAS regularly fail to perform as intended in these challenging environments. Terrain classification may improve the performance of ADAS. A lot of research has been conducted on on-road terrain classification, but few studies focus on off-road terrain classification. This study classifies several off-road terrains, based on road roughness using the ISO8608:2016 standard, using a convolutional neural network (CNN). A database of images over different terrains with known road roughness was created using forward and downward facing cameras. Two different classification models were built: one is brand new and the other made use of transfer learning on pretrained model. Terrain data was captured on several on-road and off-road tracks. Results indicate that off-road terrain classification with cameras can be done with high accuracy before a vehicle drives over a specific part of a road.

**Keywords:** off-road terrain classification, road profiles, off-road vehicle dynamics, convolutional neural networks, supervised learning, image data

## 1 INTRODUCTION AND BACKGROUND

---

### 1.1 BACKGROUND

Injuries and fatalities have a significant impact on the global economy through both treatment costs for the injured and the loss of productivity of those killed or disabled. The World Health Organisation (WHO) estimates that road traffic injuries cost countries 3% of their annual gross domestic product. The most vulnerable to road traffic injuries and fatalities in our society are in low- and middle-income countries. The conditions in these countries often give rise to a prevalence of unsafe vehicles and deteriorating or non-existent road infrastructure. The WHO is the lead agency coordinating the United Nations' Decade of Action for Road Safety 2021-2030 that targets prevention of at least 50% of road traffic deaths and injuries by 2030 (World Health Organisation, 2020).

One of the approaches to reduce the impact of road accidents rate is the development of Advanced Driver Assist Systems (ADAS). ADAS, such as electronic stability control, automatic emergency braking, lane keep assist and adaptive cruise control, are attempts to eliminate human error, thus improving vehicle safety. Many of these high-end features are becoming more affordable and hence penetrating the lower- and middle-income markets (World Health Organisation, 2017). However, current systems are almost exclusively designed for high-income countries with properly maintained infrastructure. The effective application of these systems is often limited to ideal conditions (Velodyne Lidar Inc., 2019, Euro NCAP, 2015, Euro NCAP, 2018). This means that the application of modern safety products to challenging driving environments (as typically encountered in low- and middle-income countries) is

limited, even though this is the portion of the global society that will benefit the most from these modern safety features.

Low- and middle-income countries typically rely on agriculture and mining to generate income, meaning that the economically active portion of the population is often located in rural areas with poorly designed roads and other infrastructure (The World Bank, 2014, Ericsson and Löf, 2018). The limited application of modern safety technology to challenging driving environments needs to be addressed to maximize the benefits of these systems. Adapting intelligent safety systems to the terrains the vehicle is traversing is one of the first steps to addressing this gap.

## 1.2 TERRAIN CLASSIFICATION

Terrain classification using on-board sensors can provide important information to improve safety, fuel efficiency and passenger comfort. The vast majority of terrain classification work has focused on on-road terrain classification, not off-road terrain classification. A wide range of applications including defence, agriculture, conservation, and search and rescue could benefit from off-road terrain classification (Shaban et al., 2021).

Terrain classification is categorised as vision-based, reaction-based or a combination of reaction- and vision-based methods (Coyle, 2010). The purpose of terrain classification is to classify the terrain according to pre-determined classes and then to adapt either the information passed through to the driver, or the configuration of the vehicle's subsystems (such as suspension settings, electronic stability control algorithm, or simply to limit vehicle speed) to enhance performance and safety.

Table 1 summarizes the approaches found in the literature with the significant variations mentioned.

*Table 1: Summary of classification approaches found in the literature*

Reference	Classifier details	Sensors	Dataset specifics	Results
Selvathai et al. (2017)	Neural network with supervised learning. Binary classifier (on-road and off-road)	Camera	6250 images, 50x50p	93%
Sung et al. (2010)	Neural network with supervised learning. Multiple classes (soil, gravel, foliage and sky)	Camera	200 images per class, 16x16p	81.8%
Lu et al. (2009)	Probabilistic neural network with supervised learning. Multiple classes (asphalt, grass, gravel, sand, white sand, red sand)	Camera image of line laser projected on terrain	900 images, 640x480p	>86%
Seraji & Howard (2001)	Extraction of features (Slope, roughness, discontinuity and harshness) using different methods to classify traversability using fuzzy logic	Camera	No information provided	No information provided
Roychowdhury et al. (2019)	Convolutional neural network with supervised learning. Multiple classes (dry, wet/water, slush and snow/ice)	Camera	3750 training images and 1550 testing images	97%

Reference	Classifier details	Sensors	Dataset specifics	Results
Weiss et al. (2007)	Comparison of support vector machines, neural network, Bayes classifier, k-Nearest Neighbour Multiple classes (indoor floor, asphalt, gravel, grass, paving clay)	Accelerometer	10225 samples of 1s (100Hz) sampled vertical acceleration	87.33%
Omer and Fu (2010)	Support vector machine applied to extracted image features supervised learning. Classified between winter profiles that included bare, snow covered and track roads.	Camera	200 training images and 90 test images 160x400p	84 to 86%
Byl & Filitchkin (2012)	Support vector machine with supervised learning. Multiple classes (asphalt, grass, gravel, mud, soil and woodchips)	Camera	54 images total 320x320p	100%
Manduchi et al. (2005)	Terrain classification based on colour of pixels using maximum likelihood Gaussian mixture model. Terrain classification using Lidar and range measurements	Camera and Lidar	1313 Lidar scans, no information on camera	99%
Stolee and Wang (2018)	Different classification methods used (logistic regression, K-NN, decision trees, Mixture of Gaussian, NN) applied to 1) average RGB value, positional difference of pixels in superpixel and 2) average hue saturation value (HSV) variance in RGB, average grayscale entropy and frequency of edge pixels within superpixel to classify pixels as tarred road or not.	Camera	289 images	90.56%
Giese et al. (2017)	Course estimation of ego lanes based on occupancy grids generated by radar sensors.	Radar (operates at 76 GHz, 200 m range)	3.2 million grids with edge lengths of 0.5 m.	No information provided.
Vulpi et al. (2021)	LSTM, SVM and CNN applied to spectrogram different sensors. Multiple classes: Concrete, Dirt, ploughed and unploughed	Tri-axial accelerometer, gyroscope, left and right wheel seeds and motor current (10 channels in total)	Not specified, contains over 156 (5sec) recordings of which multiple samples are drawn	CNN 91.5%
Goto and Ishigami (2021)	CNN applied to RGB and Infra Red (IR) images of different soils with different moisture content Multiple classes: Perlite, Pumice, Leaf Mold and Dark Soil	Camera	600 images of each class 227x227p	RGB 100% IR 99.8%

Many approaches listed in Table 1 make use of reactive measurements which make use of sensor measurements while driving over the terrain (Weiss et al., 2007, Vulpi et al., 2021). These methods delay the classification of the terrain which could result in initial poor results until classification is performed. Camera based methods have the advantage that the terrain in front of the vehicle can be

classified, thus providing some preview on the terrain (Selvathai et al., 2017, Seraji & Howard, 2001, Roychowdhury et al., 2019). Convolutional Neural Networks have been used in multiple state-of-the-art classifiers (Tan and Le, 2021), in different benchmark studies such as the CIFAR dataset (Krizhevsky et al., 2009), and the Imagenet dataset (Deng et al., 2009). CNNs have been used in the classification of terrain using spectrograph images (Vulpi et al., 2021) and Red-Green-Blue (RGB) and Infra-Red (IR) images (Goto and Ishigami, 2021). Thus, the CNN architecture presents the best state of the art method in classifying images and is therefore used in this study.

This study will use camera images that can be used to classify terrain before being traversed by the vehicle. Several studies used the images in the red, green and blue (RGB) colour format, whereas other studies used the images in the greyscale or hue saturation value (HSV) format. More involved approaches separate the brightness level from the image to create robustness under differing light conditions (Byl & Filitchkin, 2012). Weiss et al. (2007) extracted integral invariant features from the images. Some studies have looked at comparing different colour spaces. Nishi et al. (2017) found HSV yield a very small improvement over RGB (97% vs 93%) in classifying fruit, whereas (Sachin et al., 2017) found RGB significantly outperformed (90% vs 78%) HSV in classifying scenes. Thus, the choice of colour space appears to be depended on the objects being classified. Since RGB is widely used in many CNNs and is the direct output of most cameras the input to the classifier will be RGB images.

All the studies made use of multiple class classification except for one which merely distinguished between on-road and off-road terrains. Each study consisted of arbitrarily selected classes that did not make use of any road profile classification standard. For ADAS on off-road terrain several terrain aspects may be useful to know. Firstly, the type of terrain (gravel, tarred, soil etc) as this generally affects mobility and longitudinal performance of the vehicle, this has thus-far been the focus of most studies although classes are often arbitrarily chosen. Another aspect is the vertical excitation induced by the terrain which has a major effect on the ride-comfort but also, due to overall loss of grip, the longitudinal and lateral performance of the vehicle as well. Both aspects are relevant information to improve off-road performance and safety. To the best knowledge of the authors no study on the classification of the road roughness using terrain images has been performed. Classification of road roughness are commonly performed using ISO8608:2016 standard (International Organisation of Standardisation, 2016).

Based on the literature survey it was decided to use a camera sensor that used a convolutional neural network (CNN) coupled with a supervised learning model. The specific contribution of this study is:

1. The creation of an image database of off-road and on-road terrain of different road roughness of terrains on a vehicle testing facility
2. The classification of terrains based on road roughness using the ISO8608:2016 standard.
3. Comparing downward and forward-facing cameras' classification performance
4. Comparing transfer learning of pretrained model on separate classification tasks and training from scratch of CNN networks

## 2 METHODOLOGY

---

The methodology followed consists of three steps:

1. Investigating terrain classification based on road roughness.
2. Data capturing and class selection making use of ISO 8608:2016 (International Organisation of Standardisation, 2016).

3. Development of classification model.
4. Evaluating the performance of the developed classifiers.

## 2.1 ROAD ROUGHNESS CLASSIFICATION

Off-road terrains are characterised by the presence of road undulations and varying surface conditions. The undulations give rise to vertical excitation of the vehicle, resulting in varying tyre normal forces. The presence of organic material, loose stones and other objects on the terrain may result in varying friction coefficients. The varying tyre normal force and available friction coefficient have detrimental effects on the performance of intelligent safety systems (Hamersma and Els, 2014).

Road classification focuses on the road roughness. Roughness is described by the elevation profile along the wheel tracks over which the vehicle passes. A road profile is typically a random signal and is described according to its statistical properties (Gillespie, 1992). ISO 8608:2016 describes the road as a displacement power spectral density (PSD) vs. spatial frequency and classifies the road according to its roughness. The roughness is determined at a reference spatial frequency (typically 0.1 cycles/m) and is denoted by  $G_d(n_0)$ . Eight classes, from A to H, are defined. A larger value of  $G_d(n_0)$  denotes a rougher road. The general form of the displacement PSD is given as:

$$G_d(n) = G_d(n_0) \cdot (n/n_0)^{-w} \quad (1)$$

Where  $G_d(\cdot)$  is the displacement PSD (in  $m^3$ ),  $n$  is the spatial frequency (in cycles/m),  $n_0$  is the reference spatial frequency and  $w$  is the exponent of the fitted PSD, typically a constant value of 2. The spatial frequency typically lies within the range of 0.01 cycles/m to 10 cycles/m (thus significantly smaller than a typical contact patch of a passenger car).

One of the shortcomings identified in the literature was the use of arbitrarily selected classes. Classes such as 'grass', 'gravel', 'snow', etc. were found in the literature. The link between these arbitrarily chosen classes and their effect on vehicle dynamics is unclear, thus limiting their value to ADAS designers. The benefit of using a well-established, standardised method of road profile classification is that ADAS designers can perform extensive simulations on random road profiles and critically evaluate the performance of their control algorithms on the different road types.

## 2.2 DATA CAPTURING OF OFF-ROAD TERRAIN CLASSES

Data capturing was conducted at the Gerotek Test Facilities in South Africa (Armcor Defence Institutes SOC). Gerotek Test Facilities was originally designed for the performance evaluation of military vehicles. It houses several off-road tracks that were sprayed with concrete to preserve the natural off-road terrain as would typically be encountered in the Southern African war theatre. The facility also contains more conventional flat asphalt and concrete terrains that represent typical on-road conditions. A good mixture of on-road and off-road surfaces were thus available to this study. Becker and Els (2014) used a road profilometer to measure the 3D road profiles of all the tracks at Gerotek Test Facilities – this data was available to this study (see Figure 1). The tracks were classified according to ISO 8608:2016 with the data recorded by a road profilometer and shown in Table 2. These classes were then used as labels for the various tracks. The classifier thus had three different classes namely a class B, D and H.

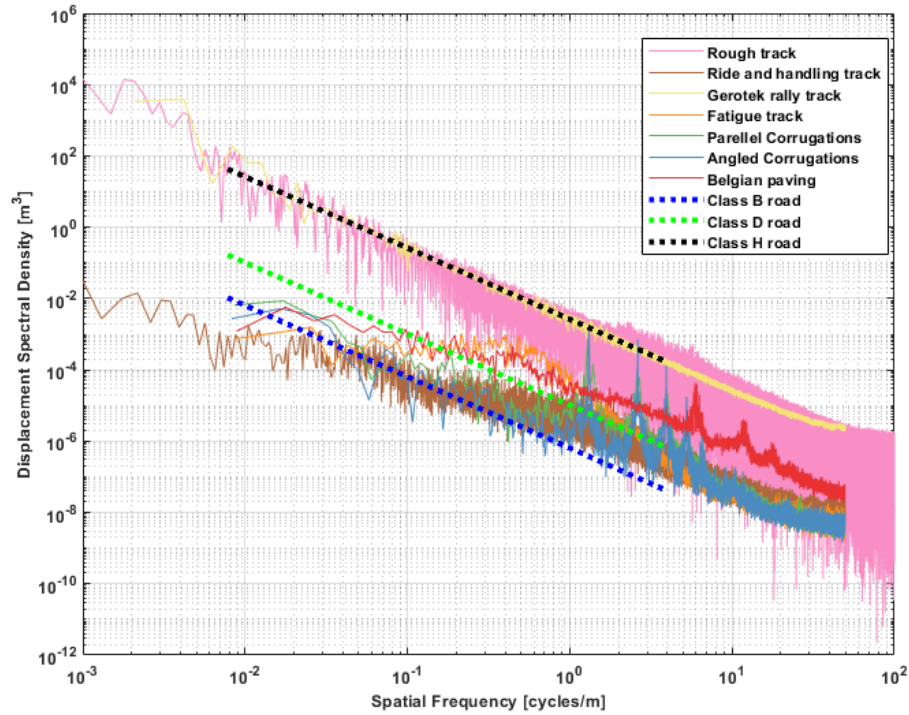


Figure 1: Displacement PSD results obtained from Can-Can data captured at Gerotek Test Facilities (Becker and Els, 2014)

Table 2: List of tracks and their classification used for this study

Track	ISO 8608:2016 classification
Gerotek rally track (gravel track)	Class H
Rough track	Class H
Belgian paving	Class D
Parallel corrugations	Class D
Angled corrugations	Class D
Fatigue track	Class D
Ride and handling track	Class B

### 2.2.1 Experimental setup and procedure

Basler Dart 1-megapixel (MP) colour cameras, with 6mm focal length lenses, were mounted on an off-road test vehicle. A comparison between two different camera orientations is made. The first orientation is a downward-facing camera in front of the vehicle. The second orientation is forward facing-camera (between 5° and 15° from the horizontal). While the downward-facing camera will provide a much shorter to no preview (dependent on mounting location) of the terrain compared to the forward-facing camera, it captures a top-down view of mostly the terrain and thus less affected by vegetation. It may be better at detecting higher frequency undulations. The forward-facing camera captures an overall larger area of the terrain which may enable detection of lower frequency undulations but is affected by loss of resolution the further away the terrain is from the camera.

Figure 2 on the left shows the forward-facing camera placed on the vehicle bull bar. The forward-facing camera was approximately 1 160 mm from the ground (attached at bonnet height). Figure 2 on the right shows the downward-facing camera mounted just above the test vehicle number plate. The camera was placed approximately 630 mm from the ground.



Figure 2: LEFT) Forward-facing camera setup on test vehicle RIGHT) Downward-facing camera setup on test vehicle

The off-road test vehicle was driven at a low speed never exceeding 30km/h across the seven tracks, trying to maintain a constant speed. Note that the very rough tracks were extremely challenging to traverse at a constant speed. The total dataset for the downward- and forward-facing categories contained 15 308 images. The total number of images per class is summarised in Table 3. Figure 3 and Figure 4 shows an example image of the different off-road terrains captured using the forward and downward facing cameras respectively.

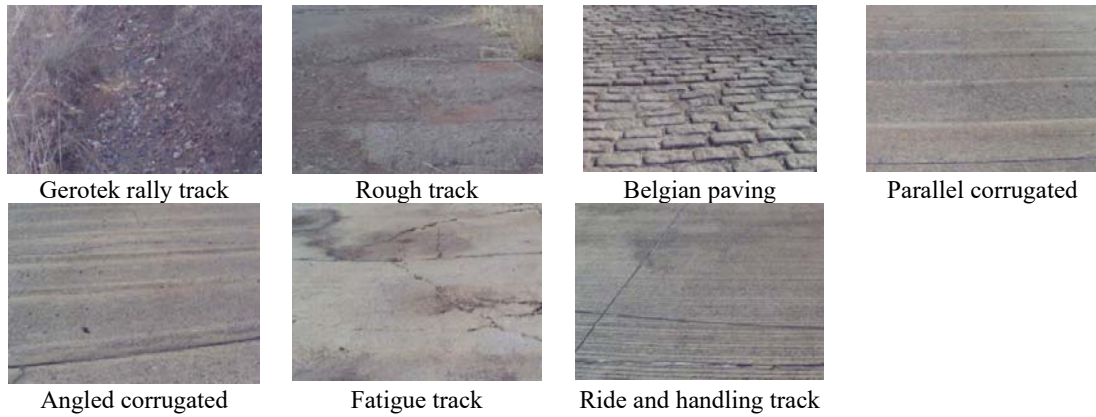


Figure 3: Example images of the different off-road terrains for forward facing dataset.



Figure 4: Example images of the different off-road terrains for downward facing dataset.

### 2.2.2 Data pre-processing

Before the classifier could classify between the different off-road terrains, the data required pre-processing first. Pre-processing varied between the downward- and forward-facing datasets.

Image labels were automatically assigned to each image based on the overall road roughness of the terrain traversed while capturing images. The image labels are the road roughness of each terrain as in Table 2. The forward-facing camera required additional cropping of the image to remove the noise



from the image (vegetation, sky etc). This process was automatically performed by selecting a fixed region in the image where the road was situated. Since the region is fixed, the region may contain noise of tree branches or road side during sharp cornering. The original captured images were  $1280 \times 960$  pixels. After removing the noise in the forward-facing dataset, the images reduced to  $900 \times 900$  pixels. Both the forward-facing and downward-facing dataset image sizes were resized (not cropped) to  $150 \times 150$  pixels and given as input for the first convolutional layer.

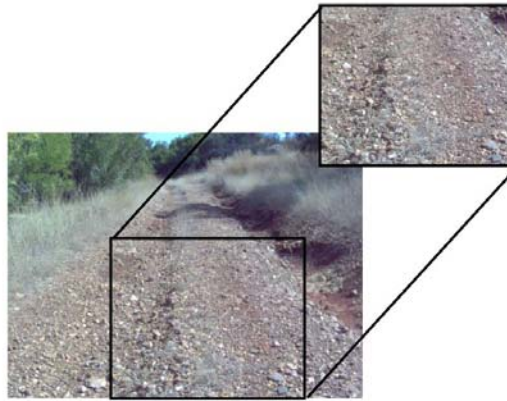


Figure 5: Example of a cropped image to omit unwanted noise

## 2.3 CLASSIFICATION MODELS

Two CNN models were developed for training: the first CNN model was created from scratch, and the second CNN model used a pretrained CNN where transfer learning was used (a feature extracting pretrained model)(Brownlee, 2019). A multiple-class classifier model was created, containing classes B, D and H. The Keras application programming interface was used, which functions on top of TensorFlow.

The architecture of the created CNN takes in  $150 \times 150$  pixel three channel RGB images. The CNN consisted of three convolutional layers with 32, 64 and 128 filters each with ReLU activation and max pooling after each layer. The final extracted features are passed through two dense layers, the first is a hidden layer of 512 nodes using a ReLU activation function and the final layer uses a linear activation function that provides the final classes of the CNN. The softmax operator was applied to the final output to transform the output into probabilities. Figure 6 shows the model hierarchy for the multiple-class classifier.

A pretrained model is used with transferring learning to also classify the terrains. Transfer learning is often used when a small database is available to train on. Since the convolution layers are trained on larger databases to classify objects this often results in much more sophisticated feature detectors in the convolution layers. The final dense classification layers can then be retrained to classify new objects. For the pre-trained model the Xception model was used (Tsang, 2018). The Xception model uses depth wise separable convolution and was 1st Runner Up in the 2015 ImageNet Large Scale Visual Recognition Challenge (ILSVRC 2015) and freely downloadable. During training the convolutional layers were frozen (not updated during training) and the dense classification layers were the only layers updated during training.

During training of both models, dropout layers were used to prevent overfitting. The AdamOptimizer and cross-entropy loss function was used (Bashaev, 2018).



## Multiple Class Classifier Model Hierarchy

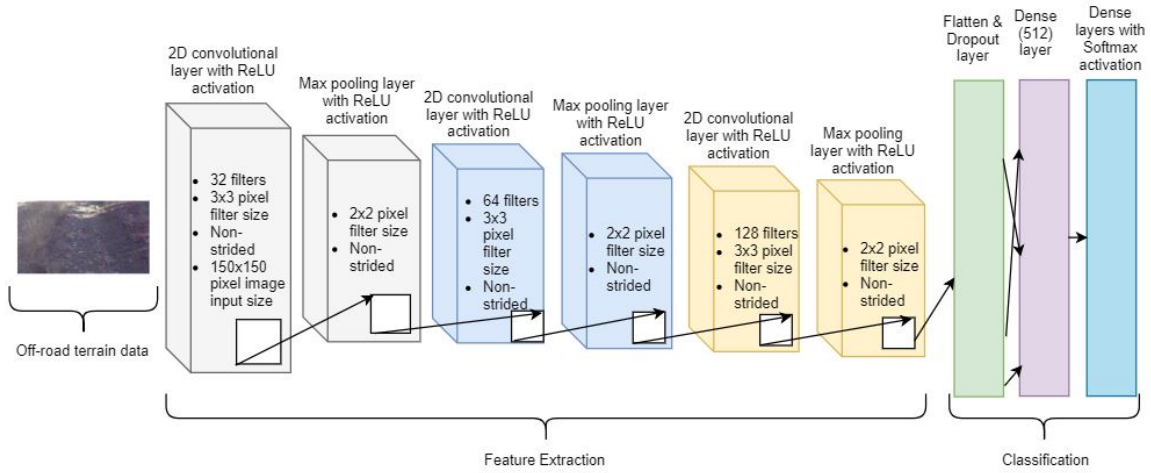


Figure 6: Multiple-class classifier model hierarchy

## 3 RESULTS

A total number of 15 308 images were captured during the data capturing phase: 7 783 images for the forward-facing dataset and 7 525 images for the downward-facing dataset. A test train split of 70/30% was used. Some classes had more images than others, thus the class with the smallest dataset was chosen to be the limit of the number of images allowed per class, as shown in Table 3. Table 3 lists the total number of images captured per class as well as the reduced total (12 645 images). This meant that 4 095 training images and 1 755 testing images were used for the downward-facing setup, whereas 4 756 training images and 2 039 testing images were used for the forward-facing setup.

Table 3: Number of images per class and total images vs reduced total images used during training

Road class	Downward facing	Forward facing
B	1 950	2 265
D	2 713	2 751
H	2 862	2 767
<b>Total</b>	<b>7 525</b>	<b>7 783</b>
<b>Reduced total</b>	<b>5 850</b>	<b>6 795</b>

### 3.1 FORWARD-FACING RESULTS USING MULTIPLE-CLASS CLASSIFIER

The model was trained on 1 000 epochs with a batch size of 128. After 1 000 epochs, the model showed signs of overfitting and thus training was stopped. The batch size is the number of samples processed before the model is updated, while the epochs used is the number of complete passes through the training dataset.

Figure 7 (a) shows both the training and the validation accuracy against the number of epochs. A training accuracy of 100% was achieved after 1 000 epochs. The validation accuracy ranged between 99.3% and 100%. At the end of the 1 000 epoch values, the validation accuracy started to decrease slightly, which was an indication of the model overfitting.

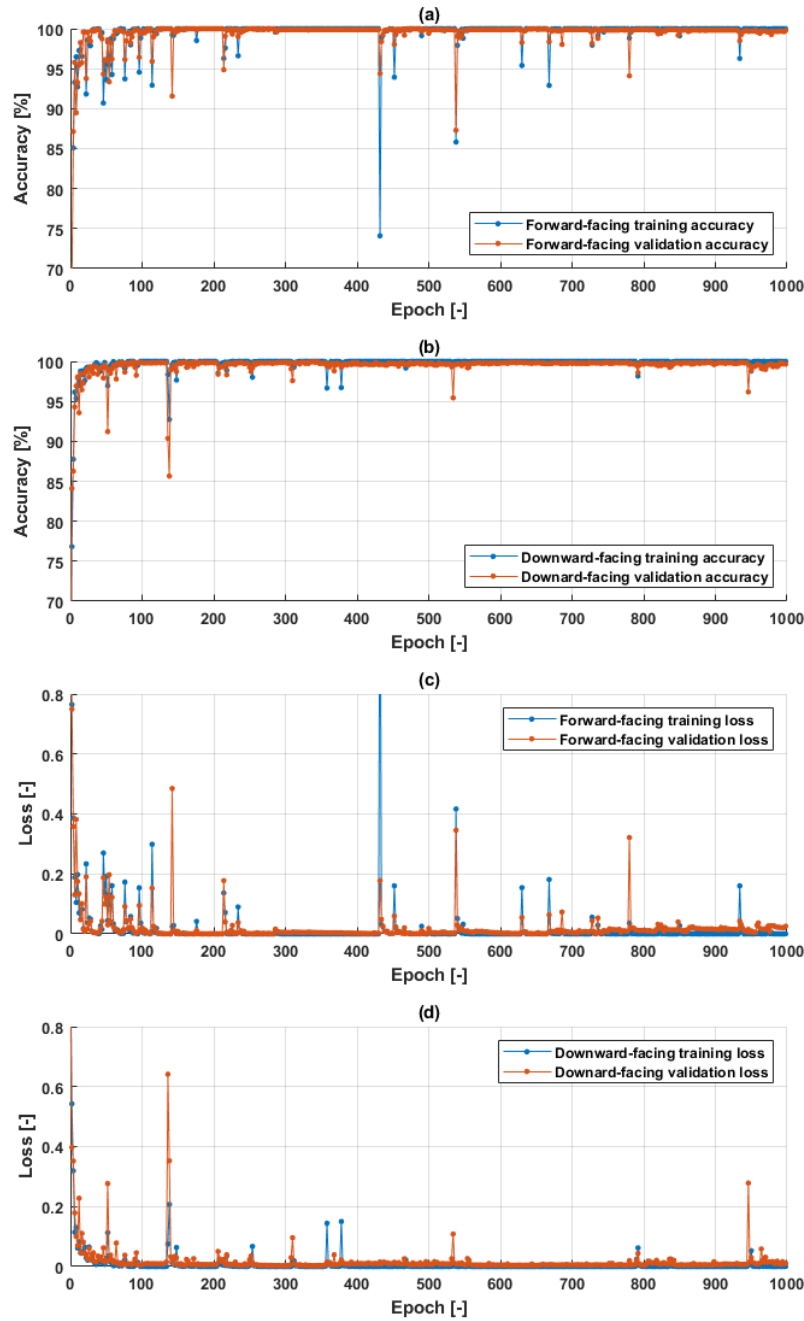


Figure 7 Training accuracy and validation accuracy results for the forward-facing dataset using multiple class classifier (a) and the downward-facing dataset using multiple class classifier (b) Training and validation loss results for the forward-facing dataset using multiple class classifier (c) and the downward-facing dataset using multiple class classifier (d)

The training and validation loss values are shown Figure 7 (c), which is a representation of whether enough epoch values were used during the training process. As the training and validation loss results approached 0, it indicated that the model had trained for enough epoch values. The results showed that the validation loss was slightly larger than the training loss at the end. This meant that the model had started to overfit. The data was clipped to exclude a potential outlier at 431 epochs, as seen in Figure 7 (c). When the data was clipped it became clear that overfitting started to occur after 600 epochs. The validation loss value remained constant between 300 and 600 epoch values.

### 3.2 DOWNWARD-FACING RESULTS USING A MULTIPLE-CLASS CLASSIFIER

Figure 7 (b) shows both the training accuracy against the number of epochs and the validation accuracy against the number of epochs for the downward-facing results using a multiple class classifier. A training accuracy of 100% was achieved after 1 000 epochs. The validation accuracy ranged between 99% and 100%. At the end of the 1 000 epoch values, the validation accuracy also started to decrease, indicating that the model was overfitting. The goal was to train the model for as few epochs as possible while achieving the best possible accuracy. Therefore, many epochs were used to determine at which point overfitting might occur.

Figure 7 (d) shows the training and validation loss values for the downward-facing results using the multiple class classifier. The training and validation loss values approached 0, indicating the model had trained for a large enough number of epoch values. As the epoch values approached 1 000, the results again started to indicate overfitting. The accuracy and loss plots for the downward-facing setup had fewer spikes in the results, but shared the same overall shape as the forward-facing setup. The correct number of epochs to use will vary for each application. It is important to analyse the results to find the point where the accuracy results approach 100 % or remain constant (i.e. show no further improvement). Also make sure that the loss tends to zero or remains constant with increasing epochs.

The model achieved a prediction accuracy of 100% for both the class B and class D road types. A 99% prediction accuracy was achieved for the class H road with the other 1% being predicted as a class D road. An overall prediction accuracy of 99.65% was achieved. The prediction results were 0.06% better than the forward-facing results; this is an almost negligible improvement. Changing the epoch values or other hyperparameters of the CNN could alter the results slightly. It is safer to assume that both the forward- and downward-facing setups gave satisfactory results when using the multiple-class classifier.

There were sudden drops in accuracy and loss percentages for both the forward- and downward-facing results. This could have been due to two reasons. First, the model is built on TensorFlow, which can cause a sudden drop in accuracy performance due to momentary memory loss. However, this did not affect the overall accuracy results and prediction accuracy on the prediction set. Second, the accuracy could drop due to the learning rate being too high, which was unlikely. If the learning rate is too high, the model usually does not recover and yield good accuracy results again. The accuracy results will drop suddenly and remain that low.

### 3.3 FORWARD-FACING RESULTS USING THE PRETRAINED MODEL

Figure 8 shows the results obtained with the pretrained model. A training accuracy of 100% was achieved after 1 00 epochs. The validation accuracy results ranged between 99% and 100%. Compared with the multiple class classifier, the pretrained model did not experience the gradual decrease in validation accuracy, but did have more spikes in the data. These spikes happened from time to time and caused a sudden decrease in validation and training accuracy. It was expected that the sudden drops in the validation and training accuracy were caused by the batch size. A batch size of 16 images was used compared with the batch size of 128 for the multiple class classifier. The smaller batch size was used because of the memory limitations of the computer being exceeded during training. The convergence graph would be smooth if the model trained on the entire dataset, instead of with batches.

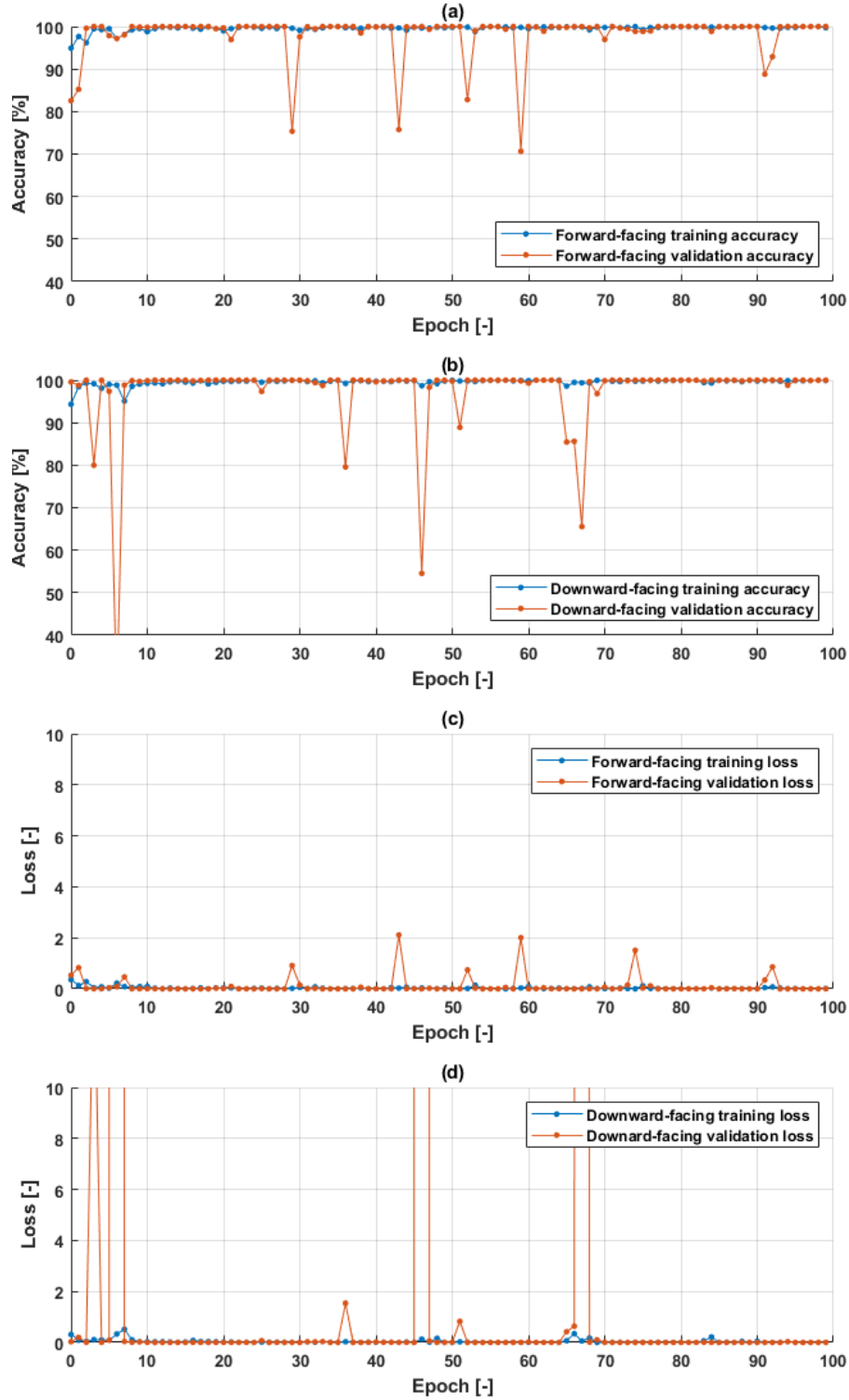


Figure 8 Training accuracy and validation accuracy results for the forward-facing dataset using pretrained model (a) and downward-facing dataset using pretrained model (b) Training and validation loss results for the forward-facing dataset using pretrained model (c) and downward-facing dataset using pretrained model (d)

The training and validation loss values are shown in Figure 8 (c) and (d). The model experienced one big spike in validation loss during the early stages of training. Although several other spikes also occurred in the data, they were not as clear due to the potential outlier. The data was clipped and the validation loss spikes can be seen easily. These spikes occurred because of the small batch size.

### 3.4 DOWNWARD-FACING RESULTS USING A PRETRAINED MODEL

A validation and training accuracy of 100% was achieved after 1 00 epochs. The validation and training accuracy again showed several spikes during the training process. The training and validation accuracies are shown in Figure 8 (b).

The training and validation loss values are shown in Figure 8 (d). The data was again clipped to remove potential outliers at epochs 6, 42 and 67 to get Figure 8 (d), which shows a clearer picture. Fewer loss spikes are seen than for the forward-facing pretrained model. A prediction accuracy of 100% was achieved for all the classes.

### 3.5 CLASSIFICATION RESULTS OMITTING CERTAIN OFF-ROAD TERRAINS

It is clear that the models are trained successfully obtaining very low training losses. However, it remains unknown if the classifier is very good at identifying a particular terrain rather than the road roughness. Therefore, an additional study was performed where one of the roads was completely omitted during training but used during testing. Would the classifier be able to predict the class of the untrained terrain correctly? The following off-road terrains were omitted from each class:

- Test 1: Parallel corrugations from class D
- Test 2: Fatigue track from class D
- Test 3: Rough track from class H

It was not possible to omit a terrain from class B as it only consisted of one terrain. The classification was performed on both the forward- and downward-facing datasets. Classification was done for both the multiple-class classifier as well as the pretrained model. After the models had been trained, the prediction model was run using two approaches. For the first approach, the validation dataset consisted of the off-road terrains used during training, referred to as method 1. The untrained terrain was added back to the corresponding class it belonged to during testing. The second approach performed the prediction task only on the untrained class to determine how many of the terrain images were classified as the correct class, referred to as method 2.

#### 3.5.1 Classification without parallel corrugated terrain from the class D road profile

Table 4 also summarises the prediction accuracy results for method 1 and method 2 for both the forward-facing and downward-facing datasets.

A 98.53% prediction accuracy was achieved for the forward-facing dataset using the multiple class classifier and a 100% using the pretrained model. A 99.32% prediction accuracy was achieved for the downward facing dataset using the multiple class classifier and a 100% using the pretrained model. The results showed a slight decrease in prediction accuracy for the multiple class classifier. The pretrained model still got a 100% prediction accuracy.

For the downward-facing dataset, only classifying the left-out terrain, a prediction accuracy of 87.97% for the forward-facing dataset using the multiple class classifier was achieved and a 100% prediction accuracy for the pretrained model. A 99.46% prediction accuracy was achieved for the downward-facing dataset using the multiple class classifier and a 100% accuracy for the pretrained model.

The prediction results were still satisfactory even though the model never trained on the parallel corrugated terrain. This could be due to the similarities between the angled and parallel corrugated terrains as shown in Figure 3 and Figure 4. These terrains share similarities, which make it easier for the model to predict the class of the untrained profile. The pretrained model results were better than the multiple-class classifier results. The prediction accuracy was 87.97% for the forward-facing multiple-class classifier results with 8.24% of the parallel corrugated terrain being classified as a class H road and 3.78% as a class B road. The prediction was done on 740 images.

### 3.5.2 Classification without fatigue track from the class D road profile

The fatigue track was now omitted because it was expected to be less correlated to the other tracks in the class D class.

A 99.56% prediction accuracy was achieved for the forward-facing dataset using the multiple class classifier and a 99.41% using the pretrained model. A 99.49% prediction accuracy was achieved for the downward-facing dataset using the multiple class classifier and a 99.66% using the pretrained model. Again, it was shown that the prediction accuracy results were very high even with an untrained track that shared fewer similarities. The prediction results did decrease, but by less than 1%.

For the downward-facing dataset, a 98.71% prediction accuracy for the forward-facing dataset using the multiple class classifier was achieved and a 100% prediction accuracy for the pretrained model. A 99.03% prediction accuracy was achieved for the downward facing dataset using the multiple class classifier and a 98.71% accuracy for the pretrained model. All the results were close to 100%. The prediction was again done on 740 fatigue track images. The forward-facing results for the multiple-class classifier delivered better results than the parallel corrugated results.

### 3.5.3 Classification without rough track from the class H road profile

A 93.66% prediction accuracy was achieved for the forward-facing dataset using the multiple class classifier and a 97.20% using the pretrained model. A 95.73% prediction accuracy was achieved for the downward facing dataset using the multiple class classifier and a 93.32% using the pretrained model. The results were all above 90% with the lowest being the multiple-class classifier prediction results for the forward-facing dataset. The results were lower than the previous sections.

For the downward-facing dataset, a 28.91% prediction accuracy for the forward-facing dataset using the multiple class classifier was achieved and a 53.59% prediction accuracy for the pretrained model. A 37.16% prediction accuracy was achieved for the downward facing dataset using the multiple class classifier and an accuracy of 83.58% for the pretrained model.

These prediction results, with exception of the downward-facing pretrained model, are not acceptable. When the model only attempted to predict the untrained profile, the prediction results were much lower than the previous sections. The pretrained models gave the best results, but only the downward-facing pretrained model had good results.

The omitted section from the rough track consists of the natural terrain that was sprayed with concrete. The rally track consists of grass, mud, rocks and gravel. The appearance of the rally track is completely different from the concrete rough track. If the untrained profile is completely different than the other terrains in its respective classes, the prediction accuracy decreases. To predict a class of an untrained profile accurately using a supervised learning model, there must be some similarities between the untrained profile and the other profiles of that class. This was the case for the parallel corrugated terrain and the fatigue track. These results indicate that currently colour plays an important role. The rally track and the rough track consist of vastly different colours. It appears that the classification model made strong use of colour to distinguish between terrains and potentially less

use of the actual roughness. There could be a correlation between colour and roughness, whereas the roughness of the terrain was paired with a certain colour scheme, but this will not always be the case. This issue could potentially be solved by using stereo-cameras that provide depth information along with colours and shadows.

Table 4: Prediction accuracy results for method 1 and method 2 for both forward facing and downward facing datasets.

		Prediction accuracy Method 1 (all terrains) [%]	Prediction accuracy Method 2 (only missing terrain) [%]
<b>Forward-facing classification – multiple class classifier</b>	<u>Without omitting classes</u>	99.71	N/A
	Omitting parallel corrugated track (class D)	98.53	87.97
	Omitting fatigue track (class D)	99.56	98.71
	Omitting rough track (class H)	93.66	28.91
<b>Forward-facing classification – pretrained model</b>	<u>Without omitting classes</u>	100	N/A
	Omitting parallel corrugated track (class D)	100	100
	Omitting fatigue track (class D)	99.41	100
	Omitting rough track (class H)	97.20	53.59
<b>Downward-facing classification – multiple class classifier</b>	<u>Without omitting classes</u>	99.65	N/A
	Omitting parallel corrugated track (class D)	99.32	99.46
	Omitting fatigue track (class D)	99.49	99.03
	Omitting rough track (class H)	95.73	37.16
<b>Downward-facing classification – Pretrained model</b>	<u>Without omitting classes</u>	100	N/A
	Omitting parallel corrugated track (class D)	100	100
	Omitting fatigue track (class D)	99.66	98.71
	Omitting rough track (class H)	99.32	83.58

## 4 DISCUSSION

The results indicate that the classification of the different roads according to the ISO8608 classes is possible. Terrains can thus be classified according to objective classes related to road roughness with the use of either a forward- or downward-facing camera along with a CNN classifier. Both the forward- and downward-facing cameras yielded similar results, indicating that some preview could be obtained using the forward-facing camera as opposed to using a downward-facing camera. Using a forward-facing camera may give the vehicle sufficient time to adapt to the upcoming terrain before it reaches it.

However, the poor result obtained when omitting the rough track during training indicates that classification may still be based on image features rather than road roughness. A much larger dataset with multiple terrains from various regions and terrain type is required to better discern whether road roughness classification is indeed possible. The inclusion of terrain profile data during training, such as obtained by Becker and Els (2014) with a profilometer, may be useful and should be seriously considered. Unfortunately, creating a database is a labour intensive and expensive process and



currently no other road classifier database with road roughness information exists. However, the result of the pretrained network and the studies where other terrains were omitted indicated that there is some possibility of the classifier generalising to road roughness and is thus worthy of further investigation.

Colour and contrast may affect classification of certain classifiers specifically if variation in these parameters were not used during training (Pawara et al., 2017, Diaz-Cely et al., 2019). Thus, it is important to capture data points for different weather conditions, as the weather conditions affect lighting and colour. For example, if it is raining, the terrain will appear completely different. Different times of the day will also make the terrains appear differently due to the position of the sun (for example, shadows and when the sun rises and sets). The dataset used was recorded on a sunny day with no rain and the data capturing occurred more or less during the midday.

NN based classifiers are also poor at extrapolating and cannot be used to classify something not trained upon. Thus, the classifier cannot classify a class C road. Ideally a class C road would be classified as either class B or D thus interpolating between the classes. This, however, cannot be guaranteed with small datasets.

These aspects necessitate the creation of a much larger database captured during different weather conditions and time of day with many more terrains of each class. Artificial augmentation can also be used to artificially increase database size and create a more robust classifier. Additional pre-processing methods could also be investigated to determine if RGB images could be processed to better provide road roughness information.

## 5 CONCLUSION AND RECOMMENDATIONS

---

This paper presented a novel classification of roads based on their roughness using a network trained CNN from scratch and the pretrained Xception CNN with transfer learning. New image databases of terrains with known road roughness using a forward and downward camera were created. High classification accuracy was obtained when trained on all roads. A subsequent study where certain terrains were completely removed during training, but then tested on, yielded mixed results indicating that there may be some potential of the classifier generalising on the image features and not on road roughness. However, results from a pretrained Xception CNN and results from other omitted terrains provide some promising results in a CNN capable of classifying roads based on their roughness.

A much larger database is required which incorporates many more roads with different road roughness and at different weather conditions and time of day. Additionally, other pre-processing methods could be evaluated for improving robustness and performance of the classifier.

Challenges associated with terrain classification found in the literature were also experienced during the study. The following challenges were specifically mentioned in the literature:

- Selvathai et al. (2017) noted that varying light conditions affected the results poorly. Seraji & Howard (2001) mentioned that illumination was a limitation due to shadowing. Similar challenges were encountered during this study. A possible workaround is the utilization of different sensor technologies (such as a camera and forward-facing lidar) and combining them intelligently.
- Lu et al. (2009) had misclassifications between the grass and gravel classes. This study addressed this issue by cropping images and thus limiting the presence of image noise such as grass and other flora in the dataset.

- Weiss et al. (2007) and Stolee and Wang (2018) both noted that the dataset size used was limiting and too small. The dataset used in this study was significantly larger than the majority of the datasets found in the literature, but an even larger dataset with more variance (such as lighting conditions, presence of image noise and more road classes) will aid in the robustness of the developed model. Nevertheless, the approach presented in this study may be used to train a terrain classifier on a larger dataset. Time and cost implications of gathering such a dataset may, however, be prohibitive of such an effort.

In conclusion, although the work presented in this study indicates that terrain classification is indeed possible, several challenges still need to be overcome to improve the robustness of the existing approaches. The ability to classify terrains according to ISO8608 using a simple camera setup is seen as a building block that may lead to the development of vehicle control systems that can automatically adapt to the terrain the vehicle is traversing.

## REFERENCES

---

- ARMSCOR DEFENCE INSTITUTES SOC. *Gerotek Test Facilities* [Online]. Available: [http://www.armscordi.com/SubSites/Gerotek1/Gerotek01\\_landing.asp](http://www.armscordi.com/SubSites/Gerotek1/Gerotek01_landing.asp) [Accessed 12 July 2016].
- BECKER, C. M. & ELS, P. S. 2014. Profiling of rough terrain. *International Journal of Vehicle Design*, 64, 240-261.
- BROWNLIE, J. 2019. Transfer Learning in Keras with Computer Vision Models. *Machine Learning Mastery*.
- BUSHAEV, V. 2018. Adam - latest trends in deep learning optimization. *Towards Data Science*.
- BYL & FILITCHKIN, P. F. A. K. B. 2012. Feature-Based Terrain Classification For LittleDog. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 6.
- COYLE, E. 2010. Fundamentals and Methods of Terrain Classification Using Proprioceptive Sensors. Florida: Florida State University Libraries.
- DENG, J., DONG, W., SOCHER, R., LI, L.-J., LI, K. & FEI-FEI, L. Imagenet: A large-scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition, 2009. Ieee, 248-255.
- DIAZ-CELY, J., ARCE-LOPERA, C., MENA, J. C. & QUINTERO, L. The effect of color channel representations on the transferability of convolutional neural networks. Science and information Conference, 2019. Springer, 27-38.
- ERICSSON, M. & LÖF, O. 2018. Mining's Contribution to Low- and Middle-income Economies. *Extractive Industries*. Oxford: Oxford University Press.
- EURO NCAP. 2015. *Articulated Pedestrian Target Specifications* [Online]. Available: [https://www.acea.be/uploads/publications/Articulated Pedestrian Target Specifications v](https://www.acea.be/uploads/publications/Articulated_Pedestrian_Target_Specifications_version_1.0.pdf)  
[ersion 1.0.pdf](https://www.acea.be/uploads/publications/Articulated_Pedestrian_Target_Specifications_version_1.0.pdf) [Accessed 22 January 2020].
- EURO NCAP. 2018. *Global Vehicle Target Specification* [Online]. Available: [https://cdn.euroncap.com/media/39159/tb-025-global-vehicle-target-specification-for-](https://cdn.euroncap.com/media/39159/tb-025-global-vehicle-target-specification-for-euro-ncap-v10.pdf)  
[euro-ncap-v10.pdf](https://cdn.euroncap.com/media/39159/tb-025-global-vehicle-target-specification-for-euro-ncap-v10.pdf) [Accessed 22 January 2020].
- GIESE, T., KLAPPSTEIN, J., DICKMANN, J. & WOHLER, C. 2017. Road course estimation using deep learning on radar data. *IEEE Xplore Digital Library*.
- GILLESPIE, T. D. 1992. *Fundamentals of vehicle dynamics*, Warrendale, PA, SAE International.
- GOTO, T. & ISHIGAMI, G. 2021. CNN-Based Terrain Classification with Moisture Content Using RGB-IR Images. *Journal of Robotics and Mechatronics*, 33, 1294-1302.
- HAMERSMA, H. A. & ELS, P. S. 2014. Improving the braking performance of a vehicle with ABS and a semi-active suspension system on a rough road. *Journal of Terramechanics*, 56, 91-101.

- INTERNATIONAL ORGANISATION OF STANDARDISATION 2016. ISO 8608:2016 Mechanical vibration - Road surface profiles - Reporting of measured data. Geneva, Switzerland.
- KRIZHEVSKY, A., NAIR, V. & HINTON, G. 2009. Cifar-10 and cifar-100 datasets. URL: <https://www.cs.toronto.edu/kriz/cifar.html>, 6, 1.
- LU, L., ORDONEZ, C., COLLINS, E. G., DUPONT, J. M. & DUPONT, E. M. 2009. Terrain Surface Classification for Autonomous Ground Vehicles Using a 2D Laser Stripe-Based Structured Light Sensor. Researchgate.
- MANDUCHI, R., CASTANO, A., MATTHIES, L. & TALUKDER, A. 2005. Obstacle Detection and Terrain Classification for Autonomous Off-Road Navigation. *Springer Science and Business Media*, 22.
- NISHI, T., KUROGI, S. & MATSUO, K. Grading fruits and vegetables using RGB-D images and convolutional neural network. 2017 IEEE Symposium Series on Computational Intelligence (SSCI), 2017. IEEE, 1-6.
- OMER, R. & FU, L. 2010. An automatic image recognition system for winter road surface condition classification. *IEEE Xplore Digital Library*.
- PAWARA, P., OKAFOR, E., SCHOMAKER, L. & WIERING, M. Data augmentation for plant classification. International conference on advanced concepts for intelligent vision systems, 2017. Springer, 615-626.
- ROYCHOWDHURY, S., ZHAO, M., WALLIN, A., OHLSSON, N. & JONASSON, M. 2019. Machine Learning Models for Road Surface and Friction Estimation using Front-Camera Images. *researchgate*.
- SACHIN, R., SOWMYA, V., GOVIND, D. & SOMAN, K. Dependency of various color and intensity planes on CNN based image classification. International symposium on signal processing and intelligent recognition systems, 2017. Springer, 167-177.
- SELVATHAI, T., VARADHAN, J. & RAMESH, S. 2017. Road and Off Road Terrain Classification for Autonomous Ground Vehicle. *International Conference on Information, Communication & Embedded Systems (ICICES 2017)*, 3.
- SERAJI & HOWARD, H. 2001. Vision-Based Terrain Characterization and Traversability Assessment. Pasadena: California Institute of Technology.
- SHABAN, A., MENG, X., LEE, J., BOOTS, B. & FOX, D. 2021. Semantic Terrain Classification for Off-Road Autonomous Driving. University of Washington.
- STOLEE, J. & WANG, Y. 2018. A Survey of Machine Learning Techniques for Road Detection. Toronto: Toronto University.
- SUNG, G. Y., KWAK, D.-M. & LYU, J. 2010. Neural Network Based Terrain Classification Using Wavelet Features. *J Intell Robot Syst*, 13.
- TAN, M. & LE, Q. Efficientnetv2: Smaller models and faster training. International Conference on Machine Learning, 2021. PMLR, 10096-10106.
- THE WORLD BANK. 2014. *For Up to 800 Million Rural Poor, a Strong World Bank Commitment to Agriculture* [Online]. Available: <https://www.worldbank.org/en/news/feature/2014/11/12/for-up-to-800-million-rural-poor-a-strong-world-bank-commitment-to-agriculture> [Accessed 21 June 2022].
- TSANG, S.-H. 2018. Review: Xception - With Depthwise Separable Convolution, Better Than Inception-v3 (Image Classification). *towardsdatascience*.
- VELODYNE LIDAR INC. 2019. *The limitations of current ADAS testing scenarios white paper* [Online]. Available: <https://velodynelidar.com/downloads/> [Accessed 13 January 2020].
- VULPI, F., MILELLA, A., MARANI, R. & REINA, G. 2021. Recurrent and convolutional neural networks for deep terrain classification by autonomous robots. *Journal of Terramechanics*, 96, 119-131.
- WEISS, FECHNER, N., STARK, M. & ZELL, A. 2007. Comparison of Different Approaches to Vibration-based Terrain Classification. *DBLP*, 6.
- WORLD HEALTH ORGANISATION. 2017. *Save LIVES - A road safety technical package* [Online]. Geneva: World Health Organisation. Available: <https://www.who.int/publications/i/item/save-lives-a-road-safety-technical-package> [Accessed 21 June 2022].

WORLD HEALTH ORGANISATION. 2020. *Road traffic injuries* [Online]. Available: [https://www.who.int/violence\\_injury\\_prevention/road\\_traffic/en/](https://www.who.int/violence_injury_prevention/road_traffic/en/) [Accessed 29 March 2020].