Neural Computing and Applications (2023) 35:14183–14194
https://doi.org/10.1007/s00521-023-08479-z

ORIGINAL ARTICLE

Road surface classification using accelerometer and speed data: evaluation of a convolutional neural network model

Ashwin Sabapathy[1] · Avik Biswas[1]

Received: 22 October 2022 / Accepted: 8 March 2023 / Published online: 23 March 2023

Abstract

Assessing the quality of road surfaces for maintenance has typically been an expensive exercise for local government agencies. This paper evaluates the feasibility of using accelerometer and speed data collected from on-board diagnostic (OBD-II) devices and labelled using the PASER system of road classification to classify road quality and identify poor stretches. An ordinal logistic model and a support vector machine (SVM) classifier were first trained on features created from raw data. The SVM and an artificial neural network (ANN) were then trained separately on the raw data parameters followed by a convolutional neural network (CNN) model. All models have comparable results in terms of overall classification accuracy on validation datasets (around 67%) but the ANN and CNN models were superior in their ability to correctly identify 'Poor' and 'Good' stretches. The CNN model had the best performance among all models evaluated. The approach presented offers a low-cost method to map poor roads and identify stretches that require more attention as an alternative to more expensive traditional methods.

Keywords On-board diagnostics (OBD-II) devices · Pavement surface Evaluation and rating (PASER) · Ordinal logistic model · Convolutional neural network model

1 Introduction

The quality of road surfaces is closely related to the safety and efficiency of a region's transportation system. In addition to affecting driving comfort, damaged or deteriorated road surfaces can result in lower vehicular fuel economy and can also lead to increased wear and tear of components such as tyres and brakes. Potholes not only generate additional vibration and stress on tyres and suspension systems, but also deform rubber leading to increasing rolling resistance and energy losses. Studies have suggested that improvements in pavement roughness, which refers to large surface unevenness, could increase fuel economy by 2–6% [1].

✉ Ashwin Sabapathy
ashwins@danlawinc.com

Avik Biswas
avikb@danlawinc.com

1 Danlaw Inc, No. 1, Primpark, Primrose Road, Richmond Town, Bangalore 560 025, India

The current practice to assess the quality of road surfaces typically uses special instrumented vehicles to measure the road roughness periodically. This involves video cameras, ground penetrating radar (GPR), laser scanners, thermal scanners, GPS equipment, etc. [e.g. 2–4]. Manual methods to report pavement surface anomalies are also employed, but these are as expensive, and time and labour intensive as using instrumented vehicles. The high costs have therefore resulted in infrequent monitoring of road surface quality. A low-cost road surface monitoring system could increase the frequency of inspection and help local governments direct attention and prioritize action for maintenance of roads and improve traffic safety.

Vision-based methods that employ cameras have also been used successfully for automated pothole detection [e.g. 5–9]. Such methods have used histogram shape-based thresholding on labelled data sets to compare textures inside a potential defect shape with the texture of the surrounding non-defect pavement to determine if the region of interest represents an actual pothole. These researchers have reported over 84% overall accuracy and over 85% precision in the detection of potholes, patches, and three different types of cracks. While this is a reasonably high level of accuracy and precision, it is unclear whether such methods can detect smaller surface defects that can cause increased noise and rolling resistance. Lighting conditions of the road can affect detection of road surface anomalies, and this is another limitation of vision-based approaches.

Vibration-based approaches using accelerometers combined with GPS positioning have also been developed for automated pothole detection [e.g. 10–12]. Standard deviation of the acceleration in the *Z*-axis (vertical) was found to be a strong indicator of road conditions [10].

Eriksson et al. [12], gathered 3-axis accelerometer data at 380 Hz from multiple vehicles and used the *Z*-axis and lateral (sideways) axis to detect potholes and other classes of surface anomalies like expansion joints, railroad crossings, manholes, etc. Since accelerometer values, when driven over a pothole or anomaly, depend on how the vehicle is driven, the approach, speed, and how the sensors are mounted, a robust set of features from the sensor observations is required to accurately detect and differentiate between such anomalies. A machine learning approach on a manually labelled data set was used to train a model to detect anomalies using peak acceleration in a 256-sample window of the *Z*-axis as a feature. Other features like the ratio of the lateral to vertical acceleration, speed to *Z*-axis acceleration, etc., were also used in the pothole detection algorithm.

Vittorio et al. [13], also used a 'vertical acceleration impulse' feature from smartphone accelerometer data and this was reported to be directly proportional to the depth of potholes. The detection rate of potholes was reported to be above 80%.

Kyriakou et al. [14], adopted a supervised machine learning approach on smartphone accelerometer and gyroscope data and trained an ANN (artificial neural network) to classify four road anomalies (no defect, transverse patches, longitudinal patches and potholes/manholes) with an overall accuracy of more than 90%.

Another study carried out by Alessandroni et al. [15] showed the influence of the vehicle speed on the vertical acceleration measured through smartphone accelerometer sensors and consequently the influence on the roughness index of the roads. Roughness is a major criterion in the assessment of road conditions according to a roughness index—usually, the International Roughness Index (IRI). Zeng et al. [16] analysed smartphone-based accelerometer data under naturalistic driving conditions and showed that vibration response measured through root-mean-square (RMS) acceleration can serve as a good indicator of pavement roughness and correlates well with IRI data. The results also showed that fewer than 12 data collecting trips are needed for most pavement sections if the application collects data at a rate of 50 Hz, while 16 trips are required at a rate of 10 Hz. The study recommended that further research needs to factor in speeds while assessing roughness. Bajic et al. [17], evaluated several machine learning techniques to classify three IRI levels (Low, Medium and High), using multiple features extracted from *Z*-axis accelerometer and speed data and showed reasonable ability to classify the IRI levels—maximum precision and recall reported was 0.67 and 0.76, respectively.

Varona et al. [18] also used smartphone accelerometer data and adopted a deep-learning approach to identify different kinds of road surfaces (asphalt, concrete, cobble and dirt) and then distinguish between potholes and other destabilizations produced by speed bumps or driver actions. Several analytical approaches were evaluated and convolutional neural networks (CNN) outperformed LSTM networks and reservoir computing in terms of accuracy.

A wavelet transformation-based filter to decompose accelerometer signals into multiple scales together with spatial filters was presented by Bello-Salau et al. [19]. Road anomalies were then detected using fixed thresholds and characterization was achieved using features extracted from the filtered wavelet coefficients. This algorithm also reported high levels of accuracy, precision and false positive rates.

Most approaches discussed above focus either on specific anomalies or on estimating roughness of road surfaces. However, road maintenance decisions are based on overall road quality that combines multiple parameters such as surface defects, deformation, cracks, patches, and potholes. This is one of the primary motivations for this study, and we attempt to develop a low-cost solution to identify poor quality road stretches using accelerometer data labelled with reference to a standardized road rating system.

Rating systems such as the Pavement Surface Evaluation and Rating (PASER) rates each road segment on a scale of 1–10 with 1 being the worst condition, and 10 being the best condition (new pavement) [20]. The ratings directly correspond to the expected remaining service life as well as appropriate maintenance activities. This paper evaluates the feasibility of using accelerometer and speed data from vehicles equipped with an on-board diagnostics (OBD-II) device with inbuilt accelerometer and GPS sensors to classify the quality of road stretches. The PASER system of road classification is used to label data collected from road stretches during naturalistic driving. The PASER system is a standard used for identifying road stretches that need repairs or reconstruction. Being able to classify road quality with this reference is therefore likely to make the approach presented in this work more acceptable for local authorities.

The approach proposed in this paper provides a low-cost technique to monitor road surface quality without the need for expensive laser and radar scanner equipment nor video-based methods. Moreover, high frequency accelerometer data are challenging to interpret for such use cases and it falls upon the technique to be able to offer higher classification accuracy. This paper, therefore, evaluates several techniques that include features derived from raw accelerometer data as well as neural networks that extract features automatically. The approach used in this paper is primarily motivated by the success of deep learning methods, and in particular, convolutional neural networks (CNN) in human activity recognition using tri-axial accelerometer data [21–23], where only a few parts of the continuous signal stream are relevant to the concept of interest—thus, local dependency and time invariant characteristics of an acceleration time series are extracted and serve to improve classification results. We hypothesize that this would be similar in context to accelerometer data collected while driving on different road surface stretches and the features extracted from a CNN-based approach would improve classification accuracy. In this paper, we evaluate a CNN model against other conventional models to classify triaxial accelerometer data into different road surface classes using a similar approach.

2 Materials and methods

The paper evaluates the classification of road surface quality using features extracted from raw accelerometer data collected from OBD-II devices as well as from neural network models with automatic feature extraction. The use of accelerometer sensors from OBDII devices ensures that only g-forces from the vehicle's movements are captured. Most smartphones are equipped with accelerometers but using such devices for road surface classification would be affected by changing the phone orientations, and handling such cases poses several challenges.

2.1 Data collection

Commercially available GPS data loggers that plug into a vehicle's OBD-II port are typically available with a 3-axis accelerometer sensor. These data loggers are powered by the vehicle's battery and are designed to poll OBD parameter data from the vehicle's CAN (Controller Area Network) bus. The data loggers can be configured to collect OBD parameters, GPS data as well as accelerometer data at different sampling intervals. This study used Gen 2.5 vehicle telematics dataloggers (Danlaw DL860, Datalogger 8 Series), which are hybrid wireless communication devices that enable data communication and connectivity via 3G UMTS/HSPA or CDMA 1XRTT, BLE wireless connections. The devices were configured to send encrypted data through the datalogger's dedicated 3G UMTS with 2GSM/GPRS fallback to a backend system. The backend system converts raw binary data from the devices to a readable format and stored in a relational database. Devices were configured to report GPS and speed every second along with 24 Hz accelerometer data. The DL860 device has an inbuilt algorithm to normalize the raw accelerometer data to take into account the specific vehicle's mounting orientation of the OBDII device [24]. The normalized accelerometer data provide g-forces with $\pm$ 2 g resolution with respect to the car coordinates—the vertical (Z-axis), lateral (Y-axis) and forward (X-axis) directions.

Data were collected from two vehicles equipped with the DL860 datalogger under naturalistic driving conditions on roads in the cities of Novi and Wixom in Michigan between October 2016 and January 2017.

2.2 Data labelling and pre-processing

2.2.1 PASER system for road surface evaluation

The geo-locations of data collected from the dataloggers were labelled using the Pavement Surface Evaluation and Rating (PASER) system available for the cities of Novi and Wixom for the year 2016 when data were collected [25]. Geofences for the roads in the two cities for which the PASER rating was available were created and each of the geo-locations (latitude-longitude) from the dataloggers were checked to see if they fell within these geofences. The corresponding PASER rating from 1 to 10 was then assigned to that geo-location. The geofences were created with a 20 m buffer on either side of each road stretch to factor in GPS drift. The PASER ratings are related to needed maintenance or repair and are summarized in Table 1.

In this study, the 10 ratings of road quality were reclassified into three classes representing 'Good' (ratings 6–10), 'Fair' (ratings 4 and 5) and 'Poor' (ratings 1–3) roads. In all, data were available for 4935 s of 'Good' roads, 36,329 s of 'Fair' roads and 13,583 s of 'Poor' roads with valid data.

2.2.2 Road surface classification

In this study, road classes are represented as:

$$C = \{C_i\}_{i=1}^{m} \tag{1}$$

where 'm' denotes the number of latent road classes, i.e. three classes of 'Good', 'Fair' and 'Poor' in this work, with a corresponding sequence of accelerometer sensor readings 'a' that captures these latent classes:

Table 1 PASER rating system

PASER surface rating	Visible distress	General condition/treatment measures	Maintenance/repair required
10—Excellent	None	New construction	No maintenance required
9—Excellent	None	Recent overlay. Like new	No maintenance required
8—Very Good	No longitudinal cracks except reflection of paving joints Occasional transverse cracks, widely spaced. All cracks sealed or tight (open less than 1/4″)	Recent sealcoat or new cold mix. Little or no maintenance required	Little or no maintenance
7—Good	Very slight or no ravelling, surface shows some traffic wear. Longitudinal cracks (open 1/4″) due to reflection or paving joints. Transverse cracks (open 1/4″) spaced 10’ or more apart, little or slight crack ravelling. No patching or very few patches in excellent condition	First signs of ageing. Maintain with routine crack filling	Routine maintenance, cracksealing and minor patching
6—Good	Slight ravelling (loss of fines) and traffic wear. Longitudinal cracks (open 1/4″–1/2″), some spaced less than 10’. First sign of block cracking. Slight to moderate flushing or polishing. Occasional patching in good condition	Shows signs of ageing. Sound structural condition. Could extend life with sealcoat	Preservative treatments (sealcoating)
5—Fair	Moderate to severe ravelling (loss of fine and coarse aggregate). Longitudinal and transverse cracks (open 1/ 2″) show first signs of slight ravelling and secondary cracks. First signs of longitudinal cracks near pavement edge. Block cracking up to 50% of surface	Surface ageing. Sound structural condition. Needs sealcoat or thin non-structural overlay (less than 2″)	Preservative treatments (sealcoating)
4—Fair	Severe surface ravelling. Multiple longitudinal and transverse cracking with slight ravelling. Longitudinal cracking in wheel path. Block cracking (over 50% of surface). Patching in fair condition. Slight rutting or distortions (1/2″ deep or less)	Significant ageing and first signs of need for strengthening. Would benefit from a structural overlay (2″ or more)	Structural improvement and levelling (overlay or recycling)
3—Poor	Closely spaced longitudinal and transverse cracks often showing ravelling and crack erosion. Severe block cracking. Some alligator cracking (less than 25% of surface). Patches in fair to poor condition. Moderate rutting or distortion (1″ or 2″ deep). Occasional potholes	Needs patching and repair prior to major overlay. Milling and removal of deterioration extends the life of overlay	Structural improvement and levelling (overlay or recycling)
2—Very Poor	Alligator cracking (over 25% of surface). Severe distortions (over 2″ deep); Extensive patching in poor condition; Potholes	Severe deterioration. Needs reconstruction with extensive base repair	Reconstruction
1—Failed	Severe distress with extensive loss of surface integrity	Failed. Needs total reconstruction	Reconstruction

Source: Adapted from Walker, Entine and Kummer [20]

$$a = \{\{x_1, y_1, z_1\}, \{x_2, y_2, z_2\}, \ldots \{x_t, y_t, z_t\}, \ldots \{x_n, y_n, z_n\}\} \tag{2}$$

where x_t, y_t and z_t denote the accelerometer readings at time t. The objective is to build a model 'f' to predict the road class 'C' based on sensor readings 'a':

$$\hat{C} = \left\{\hat{C}_j\right\}_{j=1}^{n} = f(a), \quad \hat{C}_j \in C \tag{3}$$

while the true road classes (ground truth) are denoted as:

$$C^* = \left\{C_j^*\right\}_{j=1}^{n}, \quad C_j^* \in C \tag{4}$$

where 'n' denotes the length of sequence and $n \geq$ m.

The goal of the classification is to fit a model f by minimizing the difference between predicted road class $\hat{C}$ and the ground truth road class C^*. Typically, a positive loss function $\mathscr{L}$ $(f(a), C^*)$ is constructed to reflect this difference. 'f' usually does not directly take 'a' as input, and it usually assumes that there is a projection function Φ that projects the sensor reading data $a_i \in a$ to a d-dimensional feature vector $\Phi(d_i) \in \mathbb{R}^{d}$. To that end, the goal turns into minimizing the loss function $\mathscr{L}(f(\Phi(d_i)), C^*)$.

Many machine learning techniques for time series classification have been explored in the literature. The performance of these methods depends on the types of features extracted, which in turn needs specific domain knowledge and experience. Based on the features used in past studies, features from the normalized accelerometer data (oriented with respect to vehicle coordinates) were extracted for 1-s time windows. The 1-s window was used because vehicle speeds and GPS locations were available at this frequency and features extracted from 24 Hz accelerometer data would correspond to this. Other studies have also reported feature extraction with comparable time windows. The main features from accelerometer data extracted were the mean, range, standard deviation, root-mean-square for each of the three axes. A few studies have reported that lateral *g*-forces are also impacted by poor quality roads and the resultant of the *Y* and *Z* axes (square root of the sum of squares of the *Y* and *Z* axes values) was also used to extract the same features. In addition to these features, a Fast Fourier Transformation of the time series accelerometer data was carried out to convert the series to the frequency domain. The maximum amplitude of the transformed series was also extracted as a feature.

Since speed also affects the accelerometer values, features with interaction effects of speed and individual features were also derived at the secondary level. The list of features developed and used in this study are listed in Table 2:

2.3 Classification models

2.3.1 Ordinal Logistic regression

Since the road surface quality has been labelled into three ordered classes—'Good', 'Fair' and 'Poor'—an ordinal logistic model was developed with the features aggregated from 1-s accelerometer data. Multiple models were evaluated for various combinations of features using the Log-likelihood, AIC and BIC estimates as well as significance of parameters. A model that excluded speeds less than 15 kmph was also evaluated. The pre-processing of the data consisted of scaling the variables to mean zero and unit standard deviation before creating the features. This was found to improve the model performance significantly. In standardizing the feature variables, the standardization was carried out with respect to a particular accelerometer axis, i.e. the entire set of 24 samples per second of an individual accelerometer axis in the dataset was used to determine the mean and standard deviation and applied to each value.

2.3.2 Support Vector Machines (SVM)

A SVM is a classification method that can perform with high accuracy and tend not to overfit in many cases. The objective to find a hyperplane in *n*-dimensional space with maximal separation of data points to their potential classes through linear and nonlinear kernels offers a powerful technique and is evaluated in this study to classify road surfaces. Models were first evaluated using the features developed and shown in Table 2 as input parameters. A separate exercise was also carried out with raw accelerometer data at 24 samples per second and speed at 1 sample per second as input parameters, i.e. 73 input nodes were provided as input parameters for training an SVM model to classify the road into the three classes. The raw input parameters were all standardized to the same scale before training the model.

2.3.3 Artificial Neural Network (ANN)

A feed forward linear neural network with 73 input nodes (24 samples per second of each accelerometer axis and speed) and three output nodes corresponding to the three classes was evaluated with two hidden layers of 146 nodes each. The 73 input parameters were standardized to the same scale prior to training. Adding additional hidden layers was also evaluated. A separate analysis of Fourier-transformed input variables was also carried out but the

Table 2 List of features developed

Features	Description
Mean	Mean of 24 samples of *X*-axis, *Y*-axis, *Z*-axis, resultant of *Y* and *Z* accelerometer values
Root mean square	Root mean square of 24 samples of *X*-axis, *Y*-axis, *Z*-axis, resultant of *YZ* accelerometer values
Standard deviation	Standard deviation of 24 samples of *X*-axis, *Y*-axis, *Z*-axis, resultant of *Y* and *Z* accelerometer values
Range	Range of 24 samples of *X*-axis, *Y*-axis, *Z*-axis, resultant of *Y* and *Z* accelerometer values
Maximum amplitude of FFT transformed axes	FFT transformed *X*-axis, *Y*-axis, *Z*-axis series of 24 samples
Interaction of speed with RMS	Speed and rmsX, rmsY, rmsZ, rmsYZ interaction terms
Interaction of speed and standard deviation	Speed and stdX, stdY, stdZ, stdYZ interaction terms

classification accuracy was much lower and the results are not reported here.

The features developed (shown in Table 2) were also used as input variables for an ANN model with 2 hidden layers. Here too, the features were standardized to the same scale before training the model.

2.3.4 Convolutional neural network (CNN)

Conventional approaches like decision trees, support vector machines, etc., have been quite successful for classification of sensor data. These approaches rely on heuristic hand-crafted features of raw data, although only shallow features are developed limiting performance and classification accuracy. Several studies of human activity recognition from multichannel time series sensor data have reported superior classification results using CNN, a class of machine learning models within Deep Learning, which does not rely on domain expertise to develop features, but rather extracts features from raw data through a learning process [e.g. 21, 22, 26]. The features are not explicit and are internal to the model.

CNNs or ConvNets are a specialized neural network with a special architecture composed of a sequence of convolutional, pooling, and fully connected layers that can perform classification or regression tasks. A convolutional layer abstracts the input signals to a feature with a defined shape and the outputs of this layer are passed on to the next layer. A pooling layer reduces the dimensions of data by combining the outputs of other layers, which reduces overfitting as well as speeds up the training process on large data sets. After a sequence of convolutional and pooling layers, a fully connected layer connects every neuron in one layer to every neuron in another layer before the final classification is carried out.

CNNs can extract features from signals resulting in high levels of accuracy in image classification, speech recognition and text analysis. Zeng et al. [23], treated each dimension (axis) of accelerometer data as a separate channel, similar to the RGB channel of an image, and performed convolution and pooling separately. The dimension of input data was 64 for each channel, which was then passed on to a convolutional layer (12 dimensions) and then a max-pooling layer (4 dimensions). A hidden dense layer of 1024 and then a secondary hidden layer of 20 with a final softmax classifier was used.

We have adopted a similar approach with an added channel for vehicle speed with a dimension of 24 for each channel. We evaluated several architectures beginning with a single 2D convolutional layer and increasing the number until no significant performance improvement was observed. Dropout layers were added in between these convolutional layers, and this parameter was tuned by trying out different values to minimize overfitting. In specifying the convolutional layers, filters ranging from 16, 32, 64, 128 and 256 were tried out. The outputs of the final convolutional layer were flattened and passed on to a dense layer with a softmax classifier with 3 classes.

The final architecture of the CNN model that gave the best classification included five convolutions with 16, 32, 64, 128 and 256 filters. The Relu activation function was used in each and dropout layers (with 10% drop) were added as shown in Fig. 1 to minimize overfitting. The final dense layer had 256 nodes before the softmax classifier. The input layer had all the raw data standardized to mean zero and unit standard deviation.

The entire process of data collection, labelling, data pre-processing and transformation, feature creation and modelling is summarized in Fig. 2.

3 Results

3.1 Raw data characteristics

The mean and standard deviations for the three axes and speeds are shown in Table 3. Only values for which vehicle speeds were above 15 kmph were considered as speeds below this threshold showed similar characteristics.

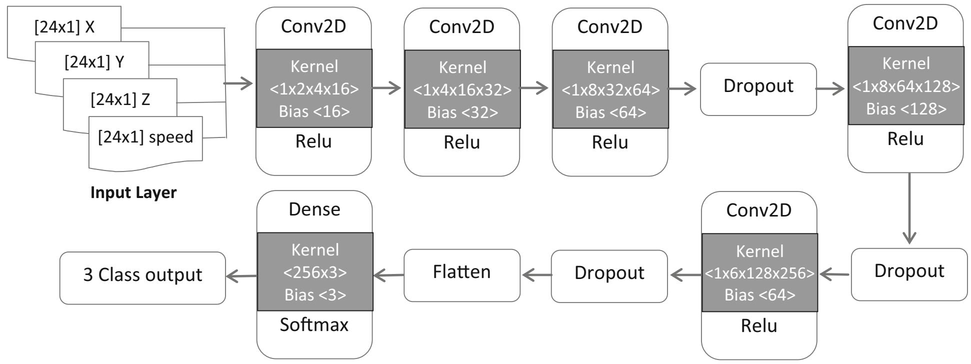
As expected, the ‘Poor’ roads have a higher standard deviation compared to the ‘Good’ and ‘Fair’ roads. The ‘Fair’ and ‘Good’ roads have comparable standard deviations. The standard deviations of the X and Y axes are higher for the ‘Good’ and ‘Fair’ roads compared to the ‘Poor’ roads, which are likely due to the higher number of manoeuvres such as braking, acceleration and lane changes on better quality roads. Poor quality roads are likely to have more uniform behaviours.

3.1.1 Feature characteristics

The mean values of the features created from raw 24 Hz accelerometer values for 1-s windows are summarized in Table 4. The summary presented is for those records with vehicle speeds above 15 kmph. Differences in mean, standard deviation, range and rms values are evident between the three classes of roads, although the differences are not very strong between the ‘Good’ and ‘Fair’ classes. The maximum amplitude of the Fourier transformed features also shows differences between the classes for the X and Y axes but not for the Z axis.

Boxplots for the mean and standard deviation features for each of the three axes are shown in Fig. 3. ‘Poor’ roads have outlier values of Z-axis standard deviations that are much higher than ‘Fair’ and ‘Good’ roads. The interquartile range of Z-axis standard deviations is also higher for

Fig. 1 CNN architecture



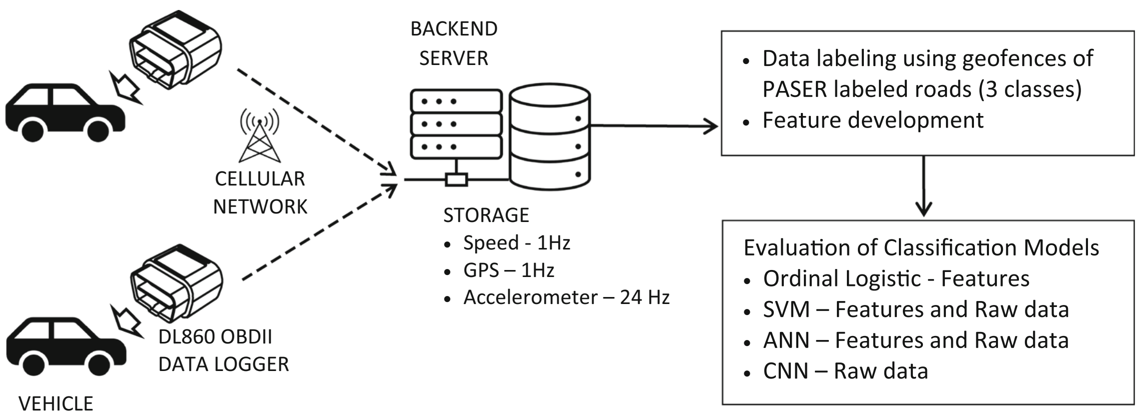


Fig. 2 Process flow of the study

Table 3 Summary characteristics of raw data by road surface quality

	Good			Fair			Poor		
	n	Mean	Std	*n*	Mean	Std	*n*	Mean	Std
X-axis	92,568	−0.008	0.071	6,96,312	−0.003	0.068	2,96,952	−0.002	0.058
Y-axis	92,568	0.000	0.064	6,96,312	0.005	0.053	2,96,952	0.006	0.035
Z-axis	92,568	0.005	0.049	6,96,312	0.995	0.045	2,96,952	0.995	0.060
Speed (kmph)	3857	55.7	15.6	29,013	57.7	15.8	12,373	61.6	15.9

‘Poor’ roads. This shows that the variability of the *Z*-axis is higher even when 1-s windows are considered. Differences are not as evident for the means and standard deviations of the 1-s means and standard deviations of the other two axes.

3.2 Model results

3.2.1 Ordinal regression

The dataset of labelled features was split through a randomized process with 70% of the data used for training the model and the remaining 30% used for predicting the classes and validating the model. The model improved when only observations with speeds above 15 kmph were considered. Interaction effects of speeds with the other features were also evaluated but these models did not show any improvement. Similarly, features extracted from the Fast Fourier Transformations also did not perform any better. The ‘stdZ’ and ‘rmsZ’ variables were highly correlated and ‘rmsZ’ was excluded from the final model. The final model summary with the variables that best explain the variance in the dependent variable is presented in Table 5.

The predicted road classes using the model on the training data set were compared with the actual road classes and the confusion matrix is presented in Table 6. The overall accuracy of the model is 67.0%. However, only 17.7% of the ‘Poor’ road class is correctly predicted and 0.7% of the ‘Good’ class is correctly predicted. 97.8% of the ‘Fair’ class is correctly predicted. The confusion matrix generated when the model is applied on the validation data

Table 4 Mean values of 1-s window-based features

	Good n = 3857	Fair n = 29,013	Poor n = 12,373
meanX	−0.008	−0.003	−0.002
stdX	0.021	0.019	0.025
rangeX	0.078	0.073	0.098
maxfftXmag	1.080	1.016	0.823
rmsX	0.052	0.049	0.045
meanY	0.000	0.005	0.006
stdY	0.022	0.021	0.025
rangeY	0.082	0.079	0.099
maxfftYmag	0.742	0.653	0.406
rmsY	0.040	0.036	0.030
meanZ	0.995	0.995	0.995
stdZ	0.044	0.039	0.053
rangeZ	0.174	0.156	0.212
maxfftZmag	23.875	23.879	23.887
rmsZ	0.996	0.996	0.997

set is also presented in Table 6. The overall accuracy achieved with the validation data is 67.3%. 18.1% of the ‘Poor’, 96.8% of the ‘Fair’ and 1.2% of the ‘Good’ classes were correctly predicted indicating no overfitting of the model.

3.2.2 SVM

The dataset used for training and validating the ordinal logistic model was also used to build an SVM classification model. Multiple kernel functions including linear, polynomial, sigmoid and radial basis function (RBF) were evaluated. There was only marginal difference in classification accuracy from among all these kernel functions evaluated. While the sigmoid kernel had the least overall accuracy of 53.8%, the proportion of correctly detected ’Poor’ and ‘Good’ roads was the highest at 38.6% and 1.8%. Applying the model on the validation data showed comparable results with overall accuracy of 53.9, and 37.9% of ‘Poor’ and 1.1% of ‘Good’ road classes detected correctly indicating no overfitting. The confusion matrices for this model are presented in Table 7.

A separate SVM model was trained using a 70% split of observations on the pre-processed raw data described earlier with the remaining 30% of observations used for validating the model. This model showed slightly lower overall accuracy of 49.3% compared to the model developed with extracted features. However, the correctly detected ‘Poor’ and ‘Good’ road classes were much higher at 35.4 and 2.4%, respectively. Here too, the sigmoid kernel performed better than the other kernels and showed no overfitting with overall accuracy of the validation data reporting 49.5% and the correctly detected ‘Poor’ and ‘Good’ classes at 33.7 and 2.8%, respectively. The confusion matrices for this model are presented in Table 8.

3.2.3 ANN

The overall accuracy of the ANN model trained on developed features as described in Sect. 2.3.3 was 70.1% with 35.2% of ‘Poor’ class roads and 10.8% of ‘Good’ class roads correctly predicted. The validation data had only slightly less overall accuracy of 68.6 and 33.8% correctly predicted ‘Poor’ class roads and 6.7% correctly predicted ‘Good’ class roads. The confusion matrices of the ANN models with developed features as inputs are provided in Table 9. The results are only reported for models with 2 hidden layers as these performed slightly better than the models with 3 hidden layers.

The ANN model trained on the raw data performed significantly better than the model trained on developed features. Although an overall accuracy of 93.4% with 88.6% of the ‘Poor’ class predicted correctly and 78.4% of the ‘Good’ class predicted correctly was achieved with this model on the training dataset, the model tends to overfit with the validation set with overall accuracy reducing to 64.1% and with 44.8 and 18.2% of ‘Poor’ and ‘Good’ classes predicted correctly. The confusion matrix of this ANN model with raw parameters as inputs is provided in Table 10.

3.2.4 CNN

The architecture of the CNN model and pre-processing of the data are as described earlier. The model was trained using the same split of 70% of observations over 50 epochs and the remaining 30% was used for validating the model. The confusion matrices for this CNN model on the training and validation data sets are presented in Table 11. There is a significant improvement in model accuracy in relation to the models presented above. The training data accuracy achieved through the CNN model is 98.9% and the correctly predicted ‘Poor’, ‘Fair’ and ‘Good’ road classes are 98.3, 99.7 and 95.8%, respectively. But when the model is applied on the validation data, the overall accuracy is 65.6%, which is comparable to the performance of other models. However, despite overfitting, the correctly predicted ‘Poor’, ‘Fair’ and ‘Good’ road classes show significant improvement with 53.5, 72.4 and 27.5%, respectively.

Fig. 3 Boxplots of Mean and Standard Deviation features

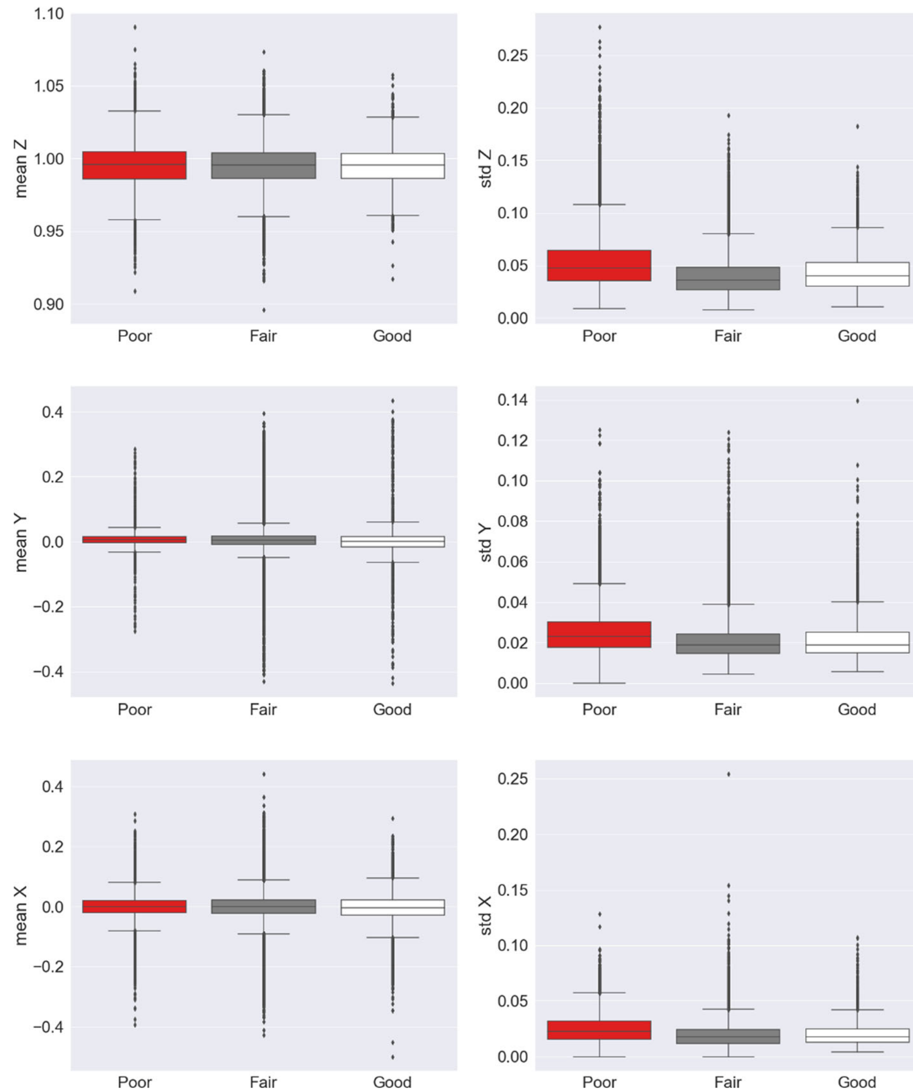


Table 5 Ordinal regression model summary

Feature	Coef	Std Error	z	$P > \|z\|$
speed	−0.082	0.014	−5.927	0.000
rmsX	0.218	0.022	10.141	0.000
stdX	−1.942	0.091	−21.383	0.000
rmsY	0.501	0.019	26.430	0.000
stdY	−1.038	0.084	−12.287	0.000
stdZ	−0.554	0.039	−14.271	0.000
Poor/fair	−2.092	0.035	−60.176	0.000
Fair/Good	1.277	0.007	193.910	0.000
No. observations: 31,670			Log-Likelihood: −25,414	
Df residuals: 31,662			AIC: 5.084e + 04	
Df model: 8			BIC: 5.091e + 04	

4 Summary and conclusions

In this paper, we evaluate the feasibility of using accelerometer and speed data collected from vehicles to classify the quality of road stretches. The accelerometer data have been collected through on-board diagnostics (OBD-II) devices equipped with accelerometer and GPS sensors plugged into the vehicles OBD-II port. The use of OBD-II devices presents a low-cost alternative to otherwise expensive solutions to monitor road surfaces. The accelerometers on OBD-II devices also do not need to be standardized as is required when accelerometers on smartphones are used—smartphones have varied makes and models of accelerometers with varying sensitivities and operational ranges and presents several challenges in pre-processing raw data.

The data collected in this work have been labelled with three classes as per the PASER system of road

Table 6 Confusion matrices—Ordinal Logistic model

		Predicted—Train data			Predicted—Validation data		
		Poor	Fair	Good	Poor	Fair	Good
True label	Poor	1536	7144	0	668	3023	2
	Fair	531	19,663	88	250	8451	30
	Good	104	2584	20	42	1093	14

Table 7 Confusion matrices—SVM model based on developed features

		Predicted—Train data			Predicted—Validation data		
		Poor	Fair	Good	Poor	Fair	Good
True label	Poor	3348	5253	79	1398	2253	42
	Fair	6265	13,658	359	2677	5892	162
	Good	862	1798	48	371	765	13

Table 8 Confusion matrices—SVM model based on raw data

		Predicted—Train data			Predicted—Validation data		
		Poor	Fair	Good	Poor	Fair	Good
True label	Poor	3077	5586	17	1245	2435	13
	Fair	7290	12,458	534	3074	5437	220
	Good	992	1652	64	433	684	32

Table 9 Confusion matrices—ANN model based on developed features

		Predicted—Train data			Predicted—Validation data		
		Poor	Fair	Good	Poor	Fair	Good
True label	Poor	3053	5593	34	1248	2422	23
	Fair	1334	18,844	104	651	7987	93
	Good	269	2147	292	115	957	77

Table 10 Confusion matrices—ANN model based on raw data

		Predicted—Train data			Predicted—Validation data		
		Poor	Fair	Good	Poor	Fair	Good
True label	Poor	7688	907	85	1653	1851	189
	Fair	296	19,774	212	1423	6,841	467
	Good	61	525	2122	169	771	209

Table 11 Confusion matrices—CNN model

		Predicted—Train data			Predicted—Validation data		
		Poor	Fair	Good	Poor	Fair	Good
True label	Poor	8539	143	9	1609	1224	176
	Fair	37	20,241	31	1909	7082	798
	Good	9	104	2557	164	398	213

classification. This labelling is important and differs from other studies in that the PASER system is a standard used for identifying road stretches that need repairs or reconstruction and is therefore likely to make the approach presented in this work more acceptable for local authorities.

We first evaluated an ordinal logistic model that was trained on engineered features built from raw data to classify road quality. This model has moderate accuracy in

Table 12 Summary of validation accuracy of models evaluated

Model	Correctly predicted			Overall accuracy (%)
	Poor (%)	Fair (%)	Good (%)	
Ordinal logistic	18.1	96.8	1.2	67.3
SVM (features)	37.9	67.5	1.1	53.9
SVM (raw data)	33.7	62.3	2.8	49.5
ANN (features)	33.8	91.5	6.7	68.6
ANN (raw data)	44.8	78.3	18.2	64.1
CNN (raw data)	53.5	72.4	27.5	65.6

both the training and validation data sets and only a low percentage of 'Poor' and 'Good' roads were correctly classified. An SVM model trained on the same features as well as trained on raw parameters also showed comparable results. An ANN model was able to achieve much higher classification accuracy in terms of correctly classifying 'Poor' and 'Good' stretches even though the overall accuracy on a validation dataset was comparable to the ordinal logistic model. The training accuracy, however, was much higher for the ANN model. We then trained a CNN model on raw data, which showed comparable overall accuracy on the validation dataset with respect to the other models evaluated, but the ability to correctly predict 'Poor' and 'Good' roads was far superior to all other models including the ANN. The results show that a CNN model with automatically extracted features performs better than the other models evaluated in identifying 'Poor' road stretches. A summary of the validation accuracy of all these models evaluated is presented in Table 12 :

While the approach presented in this paper is not a robust method for classifying roads to assist in road management and maintenance, it does offer a low-cost alternative to mapping roads and identifying stretches that require more attention using more expensive traditional methods.

This work is based on data collected from driving on asphalt roads and using the corresponding PASER rating system. We intend to widen this to concrete and other paved roads and be able to first classify the type of road followed by a classification of quality. The study also used two vehicles to collect data and we propose to have a wider variety of vehicles to see how the classification models perform.

Acknowledgements The authors gratefully acknowledge colleagues at Danlaw Inc. for their contributions to the data collection efforts. The authors are also thankful to colleagues of the Data Science team at Danlaw Inc. for assistance in data preparation and labelling.

Data availability The datasets generated during and/or analysed during the current study are not publicly available as it is proprietary to Danlaw Inc but are available from the corresponding author on reasonable request and subject to appropriate permissions from Danlaw Inc.

Declarations

Conflict of interest This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Both authors are part of Danlaw Inc. and have contributed to the work described, read and approved the contents for publication in this journal. Both authors declare no conflict of interest that may have influenced the conduct of this research and the findings.

References

1. Jackson RL, Willis JR, Arnold M, Palmer C (2011) Synthesis of the effects of pavement properties on tire rolling resistance. National Centre for Asphalt Technology (NCAT) Report, Auburn, AL, USA, pp 11–05
2. Roadscanners (2022) *Roadscanners Product Catalog* [Online]. Available at: https://www.roadscanners.com/wp-content/uploads/2017/11/RS_Catalog_ENG_web.pdf. Accessed: 18 July 2022
3. Fugro Automatic Road Analyzer (2022) *Fugro Roadware* [Online]. Available at: https://dot.alaska.gov/stwddes/asset_mgmt/assets/Fugro-ARAN.pdf (Accessed: 18 July 2022)
4. ROMDAS (2022) *ROMDAS System Overview* [Online]. Available at: https://romdas.com/downloads/documents/19-10-14%20ROMDAS%20Overview.pdf (Accessed: 18 July 2022)
5. Koch C, Brilakis I (2011) Pothole detection in asphalt pavement images. Adv Eng Inform 25(3):507–515
6. Koch C, Jog G, Brilakis I (2013) Automated pothole distress assessment using asphalt pavement video data. J Comput Civ Eng 27(4):370–378
7. Radopoulou SC, Brilakis I, Doycheva K, Koch C (2016) A framework for automated pavement condition monitoring. In: Proceedings of the construction research congress, pp 770–779, ASCE, Reston, VA
8. Radopoulou SC, Brilakis I (2016) Automated detection of multiple pavement defects. J Comput Civil Eng 31(2):04016057-1-14
9. Hadjidemetriou GM, Vela PA, Christodoulou SE (2018) Automated detection of pavement patches utilizing support vector machine classification. J Comput Civ Eng 32(1):04017073
10. Yu BX, Yu X (2006) Vibration-based system for pavement condition evaluation. In: Proceedings of the 9th international conference of applications of advanced technology in transportation (AATT), AATT, Chicago, USA, pp 183–189
11. De Zoysa KD, Keppitiyagama C, Seneviratne GP, Shihan WWAT (2007) A public transport system based sensor network for road surface condition monitoring. In: Proceedings of the 2007 workshop on networked systems for developing regions, NSDR, Kyoto, Japan, pp 1–6
12. Eriksson J, Girod L, Hull B, Newton R, Madden S, Balakrishnan H (2008) The Pothole Patrol: using a mobile sensor network for road surface monitoring. In: Proceedings of the 6th International

conference on mobile systems and applied services, MobiSys, Breckenridge, CO, USA, pp 29–39
13. Vittorio A, Rosolino V, Teresa I, Vittoria CM, Vincenzo PG, Francesco DM (2014) Automated sensing system for monitoring of road surface quality by mobile devices. Proc Soc Behav Sci 111:242–251
14. Kyriakou C, Christodoulou SE, Dimitriou L (2021) Spatial roadway condition-assessment mapping utilizing smartphones and machine learning algorithms. Transp Res Rec 2675(9):1118–1126
15. Alessandroni G, Carini A, Lattanzi E, Freschi V, Bogliolo A (2017) A study on the influence of speed on road roughness sensing: the SmartRoadSense case. Sensors 17(2):1–20
16. Zeng H, Park H, Smith BL, Parkany E (2018) Feasibility assessment of a smartphone-based application to estimate road roughness. KSCE J Civ Eng 22:3120–3129
17. Bajic M, Pour SM, Skar A, Pettinari M, Levenberg E, Alstrøm TS (2021) Road roughness estimation using machine learning. arXiv: arXiv:2107.01199
18. Varona B, Monteserin A, Teyseyre A (2019) A deep learning approach to automatic road surface monitoring and pothole detection. Pers Ubiquit Comput 24:519–534
19. Bello-Salau H, Aibinu AM, Onumanyi AJ, Onwuka EN, Dukiya JJ, Ohiz H (2018) New road anomaly detection and characterization algorithm for autonomous vehicles. Appl Comput Informat 16(1/2):223–239
20. Walker D, Entine L, Kummer S (2002) Pavement Surface Evaluation and Rating, Asphalt PASER Manual, Madison, WI: Wisconsin Transportation Information Center
21. Yang JB, Nguyen MN, San PP, Li XL, Krishnaswamy S (2015) Deep convolutional neural networks on multichannel time series for human activity recognition. In: Proceedings of the international joint conference on artificial intelligence, Buenos Aires, Argentina, pp 25–31
22. Wang J, Chen Y, Hao S, Peng X, Hu L (2017) Deep learning for sensor-based activity recognition: a survey. arXiv preprint: arXiv:1707.03502
23. Zeng M, Nguyen LT, Yu B, Mengshoel OJ, Zhu J, Wu P, Zhang J (2014) Convolutional neural networks for human activity recognition using mobile sensors. In: Proceedings of the 6th International Conference in Mobile Computing, Applications and Services (Mobi-CASE), pp 197–205. IEEE, New York
24. Greer D (2019) Vehicle monitoring module, US Patent 10,429,408 B2
25. Oakland County (2016) Pavement Surface Evaluation and Rating System, Michigan Transportation Asset Management Council. Home Interactive Map available at: https://www.mcgi.state.mi.us/tamcMap/#/
26. Alsheikh MA, Selim A, Niyato D, Doyle L, Lin S, Tan HP (2016) Deep Activity Recognition Models with Triaxial Accelerometers. The Workshops of the Thirtieth AAAI Conference on Artificial Intelligence. arXiv preprint: arXiv: 1511.04664v2