LAKEHEAD UNIVERSITY – THUNDER BAY
FACULTY OF ENGINEERING
FACULTY IN CIVIL ENGINEERING
MSc. IN CIVIL ENGINEERING PROGRAM
(2019 F) ENGI-5804-FA – MASTERS PROJECT

UNPAVED ROAD CONDITION MONITORING USING SMARTPHONES AND MACHINE LEARNING.

RAHUL TIRATH LAL (0901167)

I hereby certify that I am thoroughly familiar with the contents of this project/ assignment/ essay/report: it is substantially my own work, I have referenced all my sources of information, and I am the sole author.

SIGNATURE

NAME: RAHUL TIRATH LAL

DATE: 27/10/2019

ABSTRACT

Canada has roughly 60% (about 626,000 KM) of the public road network classified as low volume unpaved gravel roads while in the US it amounts to 53% (about 2.6 million KM) of the national road network. Poor maintenance of these roads indirectly increases vehicle operation cost by roughly 15%, also it affects the ride quality and decreases the level of service that a pavement has to offer. Therefore, it is necessary to identify these irregularities as soon as possible to allow for its timely maintenance. This was carried out through Manual condition assessment which has been a popular approach up to early 2000s. Development of technology has led to improved survey and assessment techniques such as image processing, LiDAR and using Unmanned Aerial Vehicles (UAV). Another new, innovative way to do this is through the use of sensor package available on smartphones combined with machine learning techniques. The aim of this paper is to review existing methods already available in the literature, and present a research into the use of smartphones for classification of unpaved roads utilizing the machine learning techniques including K-Nearest Neighbor (KNN) and Support Vector Machines (SVM). The results of the classifiers are also discussed along with comparisons which highlight the upper hand SVM has over other classifiers. The criteria for evaluation are Accuracy, Precision, Recall and F1-score. The appendix contains the MATLAB codes used in the process.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

According to WHO (WORLD HEALTH ORGANIZATION, 2019) report on road safety, traffic deaths continue to rise steadily reaching 1.35 million in the year 2016. It also reports that poor road infrastructure is responsible for majority traffic collisions resulting in fatal and serious injuries. Bad roads are a direct result of continuous deterioration over the design period of the pavement. It becomes very much necessary to continuously monitor pavement surface conditions to facilitate on-time maintenance. It is 3 times costlier to strengthen the pavement than to properly maintain it. (Fwa, 2005)

Development of extensive road networks in Europe and US has led to a shift from manual to semi-automated to automated data collection. Detailed manual inspections being time consuming and expensive are still extensively used contributing 99.6% of total inspection (Radopoulou and Brilakis, 2017). Due to several disadvantages of existing manual practices, the main being their expensive and time-consuming nature, research on remote sensing technologies and smartphones-based applications to judge pavement conditions are gaining more importance.

Smartphones have a number of built-in motion (accelerometers, gravity sensors, gyroscopes), environment and position sensors. Their capability to transfer data over wireless networks, high processing speed and portability render them useful especially in data collection for highway surveys. The accelerometer sensors are able to record movement along 3 axis (x, y and z-axis) and are mainly used to determine orientation of the phone (Douangphachanh and Oneyama, 2013), which can be utilized to predict the surface conditions of the pavement. The earliest research on road condition monitoring utilizing accelerometer sensors is The Pothole Patrol (Eriksson et al., 2008), which uses external hardware for data processing. It uses machine learning algorithms to analyze the accelerometer data to identify as well as differentiate potholes from other road anomalies. A false positive rate of 0.2% was achieved in controlled experiments. Microsoft also developed Nericell (Mohan et al., 2008) on an android platform to detect potholes as well as monitor traffic conditions. The detection of braking event by analyzing y-axis data gives a false negative rate of 4-11%. It uses the z-axis data to identify bumps and potholes which reported less than 10% false positives and 20-30% false negatives. Most research projects in this field were done for urban areas and paved surfaces. They have adopted machine learning techniques to improve precision in predicting road roughness using accelerometer data. Comparatively very few papers can be found in literature utilizing remote sensing or smartphones for anomaly detection in unpaved roads. This paper focuses on road condition monitoring of unpaved road using smartphones. K-nearest neighbors (KNN) and Support Vector Machine (SVM) machine learning techniques were used in this research project. This paper is divided into three sections. First is the literature review which discusses the different type of road profiles, common defects of unpaved roads, methods used for survey in the past and machine learning algorithms.

Next section describes the methodology used in this research and highlights the preprocessing and features of the selected data as well as the method of data collection. Last section presents the results and interprets and discusses the findings.

2. LITERATURE REVIEW

2.1. ROAD PROFILE

A road is a way on land between two places, improving some form of conveyance walking, motorcycle or cars. It usually consists of several layers with each layer constructed as per design guidelines. They are normally termed as paved or unpaved, further classified as flexible, rigid or composite pavements depending on the type of surface. The design process, performance and surface defects depend on the type of surface provided and is discussed in detail in this section.

2.1.1. PAVED ROADS

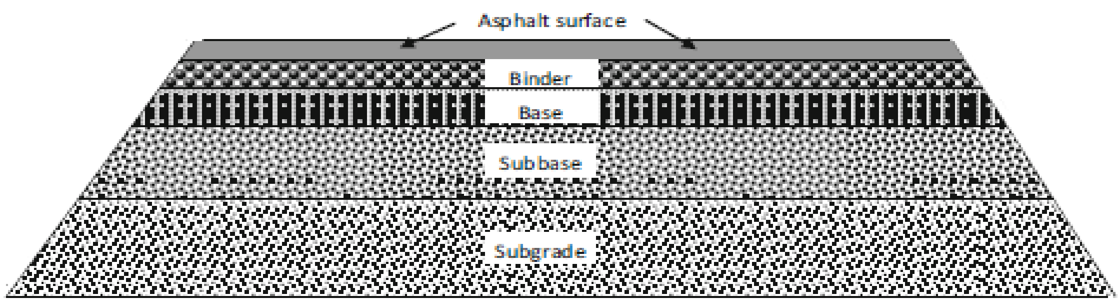

Figure 2. FLEXIBLE PAVEMENT PROFILE

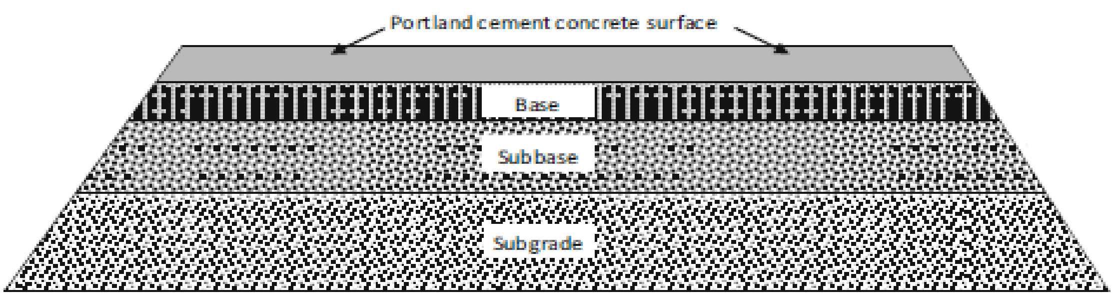

Figure 1. RIGID PAVEMENT PROFILE

Paved roads are categorized as flexible pavements (bituminous) and rigid pavements (cement concrete pavement). The major difference between the two types is the application of bituminous layer in flexible pavements and the cement concrete layer in rigid pavements as surface layers. The initial layer, subgrade comprises of stabilized or unstabilized aggregates. The top of the subgrade can be stabilized with cement or lime. A subbase course is recommended to improve the stability of the pavement. The main function of these layers is to reduce the thickness of subsequent layers. A subbase layer is 450-600 mm thick depending on properties of the subgrade. The base course serves the purpose of drainage and is made of angular aggregates with very few or no fines (Lay, 2009). It requires the use of high-quality materials tested in accordance with AASHTO and ASTM standards.

A typical flexible pavement (Figure 2) consists of surface, base and subbase courses which are built over a compacted subgrade. The surface layer is usually made of hot-mix asphalt (HMA). Localized deflections occur due to traffic load, which rebound once the loads are removed. The wheel load is distributed over a larger area as it is transferred down the layers. The thickness of each layer is determined by the magnitude and repetition of wheel loads, materials used, environmental conditions and design service life of the pavement. The thickness typically ranges from 25-250 mm including overlays. (Fwa, 2005)

Rigid pavements (Figure 1) offer high quality riding surfaces catering to daily high-speed traffic. The main material used in construction is Portland cement concrete. It is able to provide an all-weather, long-lasting surface for safe vehicular travel. Rigid pavements are able to transmit the wheel loads over a larger area. The major concerns of a rigid pavement are fatigue failure and thermally induced tensile stresses, which govern the design of the slab (size and orientation). Nominal thickness of the concrete slab is 10-30 cm, while the service life of rigid pavement is 30-40 years. Composite pavements used mainly for rehabilitation of old roads consists of a mix of both the bituminous and concrete layers. (Fwa, 2005)

2.1.2. UNPAVED ROADS

Figure 3. UNPAVED ROAD PROFILE

Unpaved roads (profile shown in Figure 3) are basically not covered with a firm level surface of asphalt or concrete. They are mainly used where the average daily traffic (ADT) is low, which does not exceed 400 vehicles. They usually act as a recreational road (e.g. to/from a ski hill resort), a resource development road or a part of the rural road systems. Hence, they are referred to as low-volume roads (LVR). They are also called dirt roads (consisting of a well compacted subgrade- fine grained soil layer) and gravel roads (existing fine-grained soil layer is reinforced with coarse grained soils mostly gravel). In more developed countries such as Canada and the US gravel roads are preferred as unpaved roads. They consist of two layers, including a well compacted soil layer (clay, silt) which is covered by coarse grained soils (gravel). A few states do not have gravel readily available and tend to use seashells, clinker, slag or recycled concrete. The fine-grained layer known as subgrade usually has naturally deposited silt or clay serving as the “foundation of the roadway” (Schnebele et al., 2015). The gradation of the gravel layer depends on the stiffness of the fine-grained soil. Larger particle size of gravel will prompt the use of fine-grained soils in conjecture to fill the voids and create a smoother surface. The gravel is filled in layers with each layer being properly wetted and compacted. (Giroud and Noiray, 1981)

Often geotextiles are used to improve to stability of the subgrade which will require a minimum cover of 15 cm from the coarse-grained material to prevent exposure and damage to it. Other properties that effect the thickness of the gravel layer are predicted annual traffic, seasonal variations and allowable serviceability loss (Selim et al., 2003). However, the thickness of the gravel layer should not exceed 1 m if the subgrade is of poor quality.

2.2. COMMON DEFECTS

Figure 4. COMMON DEFECTS: RUTTING, CORRUGATIONS, POTHOLES

Unpaved roads do not have a bituminous or concrete surface. They tend to have a wearing course of sound quality material. Therefore, surface defects are quite common compared to others. The most common examples are:

- Potholes: Normally caused by traffic action on the surface by the replacement of loose material or weak (soft) material. Potholes are bowl shaped depressions and usually considered significant if the diameter is more than 200 mm and depth is more than 25 mm. (Figure 4)
- Rutting: Presence of longitudinal depressions in the wheel-paths and caused by plastic deformations or compaction under traffic loads. Wearing courses having high clay content are more prone to rutting. (Figure 4)
- Corrugations: Regular evenly spaced transverse ridges caused by traffic action in conjunction with loose aggregate. The intervals of these ridges are normally less than 1 m. Normally occurs on hills and curves. (Figure 4)
- Gravel loss: Caused by sweeping action of traffic. It is indicative of the amount of wearing course left on the subgrade. High gravel loss causes exposure of the subgrade and requires immediate regravelling.
- Stoniness: Presence of oversized materials (excess of 40 mm) in the wearing course. It affects the riding quality and often results in tire damage due to blading action.

Another important factor to consider is erosion which is caused by movement of free-flowing water across the cross section. Erosion, stoniness, gravel loss along with loose material accumulation, are common on roads with non-cohesive fine-grained wearing courses. They can alter the cross-section increasing discomfort and safety risks. Using cohesive soils as wearing course causes slippery wet surfaces with reducing skid resistance. These defects need to be quickly identified and remedial measures such as regravelling should be adopted to rectify the cross section (Fwa, 2005). Depending on traffic and seasonal variations regravelling is to be done every 3 years, drainage maintenance every 4 years, grading twice a year and dust control once a year. Frost susceptible subgrade would require a greater thickness of gravel in re-gravelling operation. Some Canadian municipalities also look towards surface treatment as a much cheaper alternative.

2.3. PREVIOUS RESEARCH

Manual condition assessment is a conventional and reliable method for highway surveys. Pavement Inspectors either walk or ride across the sample sets of roads selected for data collection. The process is however time consuming, expensive and unsafe. The consistency of data may also differ depending on raters. Automated systems have become a part of this process due to availability of numerous sensing options such as laser sensors, video recording equipment and image processing. As of 2011, majority of US DoTs 44 out of 65 use automated pavement condition data collection systems (Pierce et al., 2013). Department of Transportation (DOT) of most states are open to vendor demonstration to implement fully automated systems as manual inspections are becoming obsolete. Remote sensing techniques utilize different regions of the electromagnetic spectrum which allows us to assess road conditions through photographs and thermal images (Schnebele et al., 2015). These are quite expensive and not feasible for use in developing countries. A cheaper alternative is presented by incorporation of smartphones in automated survey methods. Initial focus was on detection of bumps and potholes. Aksamit and Szmechta, 2011 utilized a thresholding algorithm to study the accelerometer readings and detect pavement defects. On using magnitude of total accelerometer vectors it was able to single out 6 events of unevenness from 35 runs along the selected route. Mednis et al., 2011 reported that using Z-DIFF algorithm, which measures the difference in z-axis acceleration between two points, pothole detection with 92% true positives were obtained. It was compared against threshold technique Z-THRESH, measuring standard deviation of the vertical axis, STDEV (Z)which gave lowest false positive rates when an appropriate window size had been selected (Strazdins et al., 2011) and a visual data analysis tool G-ZERO.

More recently researchers have adopted machine learning techniques to predict pavement roughness. Tai et al., 2010 used HTC Diamond mounted on a two wheeler to classify the roads in Taiwan as potholes, bumps and smooth. It was reported that the SVM algorithm achieves 78.5 % precision. A ranking system was also developed with the help of experienced professionals to estimate roughness and pavement conditions. The use of low frequency accelerometer (25 Hz) is responsible for such low accuracy. RoadSense is a real-time android application which can automatically predict the pavement quality. Tri-axial accelerometer and a gyroscope, sensors available in a Samsung Galaxy smartphone (50 Hz) were used to obtain the data. They used three algorithms namely SVM, Naïve Bayes and C4.5 Decision Tree for classifying roads into potholed or smooth. The results showed 98.6% accuracy on using the C4.5 decision tree classifier (Allouch et al., 2017). SmartRoadSense is another android application that arrives at a compound roughness index which can determine the road surface conditions. It uses Linear Predictive Coding to negate acceleration not contributing to irregularities. The Levinson- Durbin Recursion computes the prediction error for each segment of accelerometer data. Roughness index is the average of the prediction error (Alessandroni et al., 2014).

Cabral et al., 2018 focused on Paved and unpaved road classification, reporting that ResNet, a deep neural network provided an accuracy of 97% while that using SVM algorithm was 96%. For anomaly detection on paved surfaces an amalgamation of KNN-DTW performed better (100% accuracy, precision) than classical KNN and SVM (94%). Souza et al., 2018 presents an android application Asfault, to monitor asphalt pavement conditions using smartphones. To increases accuracy of prediction it uses machine learning and signal processing tools. It distinguishes between highway and urban streets depending on the speed of the vehicle and performs four-tier classification (Good, Average, Fair, Poor). The employed SVM classifier performed better than other classification techniques, (Souza et al., 2017) yielding results as high as 90% in surface classification. Yeganeh et al., 2019 validated the use of smartphones in pavement roughness estimation on basis of traveler's opinion, International Roughness Index and pavement distresses. The only negative result was achieved regarding pavement distresses as all the distresses do not affect pavement roughness. Singh et al., 2017 used DTW algorithm, used mainly for speech recognition to predict between potholes, bumps and smooth roads. It does not involve the tedious process of training the classifier and can yield good results in terms of accuracy and precision (88.99 %) at lower frequency (10 Hz) of the accelerometer. Hanson et al., 2014 did not adopt machine learning to estimate roughness. Instead the accelerometer data was double integrated to obtain pavement profile. This served as input for the software proVAL, giving us IRI values. These results were validated with profilometer readings. It was observed that at a speed of 80 km/hr the converted values were within 10% error of the actual profilometer readings. Islam et al., 2014 validated the IRI values with that of inertial profilers and obtained an average co-efficient of variability between 4 – 14% depending on test sections.

2.4. PAVEMENT ROUGHNESS MEASURE

Pavement condition is continuously monitored by government organizations in order to perform rehabilitation works at the earliest. To judge the quality of the pavement each of them are assigned indices either by an experienced rater (windshield survey) or by a high-speed profiler. Subjective evaluation by a rater is normally done to update roadway characteristics and inventory information such as pavement type, number of lanes, width, etc (Hartgen and Shufon, 1983). The high-speed profiler survey measures ride quality using the International Roughness Index (IRI), rut depth and fault height. In addition, the profiler collects forward and side view digital images of the road and right of way, and has the capability to measure radius of curve, cross-slope and grade. A few important indices used to classify pavement condition are mentioned below:

PCI (Pavement Condition Index): The deduct value based condition index developed by (LeClerc and Marshall, 1971) was further developed for the PAVER system, the oldest most commonly used distress survey method in the U.S.A. the pavement quality is expressed by a number between 0 – 100, with 0 being failed condition and 100 being excellent. Initially perfect conditions are assumed for the pavement. Each distress type and concentration has a deduct value which when subtracted from 100 will give us the Pavement Condition Index.

IRI: International Roughness Index (IRI) is a statistic calculated from a mathematical function of pavements longitudinal profile. It was developed by World Bank as a measure a roughness aimed at establishing correlation and calibration procedures. There are numerous profilers aimed at measuring IRI.

It's three categories are shown in (Sayers, 1995). Modern instruments such as profilometers and other profilers report accurate values of IRI when the vehicle operates at 80km/hr.

RQI: Ride quality index (RQI), used by many U.S agencies, is a smoothness index representing the rating a typical road user would give to the pavement's smoothness. Values of RQI ranges from 0-5. Each road is designed with a terminal RQI value which when reached indicates the need for rehabilitation procedures to pavement quality (Clyne, 2012).

RR/ RDS: Haul mine roads, a form of unpaved roads which cater mostly to heavy duty mining vehicles use rolling resistance and roughness defect score (RDS) to determine the condition of the pavement. Rolling resistance is the extra resistance to motion that the haul truck experiences. (Thompson and Visser, 2003) Roughness (RR) defect score is related to RR and is used in haul road management systems. RDS increases over time to a maximum value RDSMAX as the pavement quality deteriorates, (RR increases) implying he need for immediate maintenance. RDS depends on independent factors such as vehicle speed, wheel load, grading coefficient and shrinkage properties of the wearing course.(Thompson and Visser, 2006)

Alessandroni et al., 2014 evaluates pavement roughness in terms of Power Spectral Density (PSD). It is data expressed in terms of amplitudes and frequencies of superimposed waveforms. Roughness is defined as a PSD function of spatial acceleration, surface slope and elevation profile. Yeganeh et al., 2019 uses smartphones to gauge the ride quality index of the pavement whereas Douangphachanh and Oneyama, 2013, Hanson et al., 2014 and Islam et al., 2014 adjudge the pavement roughness in accordance to the IRI.

3. METHODOLOGY

An android application, named AndroSensor, is used in this study to record the data generated by the built-in sensors on the smartphone which are accelerometer, gyroscope, GPS, barometer, magnetometer and compass. The most important items are the vibrations recorded when the car moves on the unpaved surface (accelerometer), GPS location and time. The data are recorded at a sampling rate of 2 Hz and represented in 18 columns, accelerometer readings, linear accelerometer, gravity and gyroscope for all 3 axes x, y and z, GPS location (latitude, longitude and altitude), sound, magnetic field, orientation, atmosphere pressure, date and time. The phone was placed in a vertical position, with y-axis corresponding to the direction of gravity as seen in Figure 7. Another phone was used to record the video. The vehicle used in data collection is Ford Edge. A sedan Toyota Camry was used on a different stretch of unpaved road to compare the data recorded by the two different types of vehicles. The recording was conducted through test runs on unpaved road segments on the outskirts of the city of Thunder Bay shown in figure Figure 5, spanning approximately 6 km. Columns which were helpful in data analysis were time, location, speed and linear accelerometer readings for x, y and z-axis.

Lower speed of the vehicle at the start and end of the test run required the initial as well as the final two sets of readings to be eliminated. In addition, data collected at speeds less than 20 kmph were ignored. The time stamp on the data was matched with the video to check for discrepancies. For a few timestamps it was found that the video was not available and the subsequent data was removed. The entire dataset was then grouped into a timeframe of 3 seconds giving us a total of 212 instances. Fast Fourier Transform (FFT) was used which helped in correcting their variability (stochastic nature) by shifting the data from time domain to frequency domain. Next, the features were extracted. The machine learning algorithm performs better with data features rather than raw data. A total of 29 features shown in Table 1, 13 each of both time and frequency domain and the three axial accelerometer readings were selected. Mostly statistical features were used such as mean, median, standard deviation, peak, peak difference, kurtosis, variance, skewness, root mean square and entropy.

The instances were labelled as rough, smooth and uneven on careful analysis of the video attached, making it 91 instances with rough label, 65 smooth and remaining 56 as uneven. The profile of the accelerometer readings taken on Trendaik Road for rough, smooth and uneven pavements are shown in Figure 6. The labelling of the data is done by watching the vibrations of the vehicle with help from the video and audio data and also by carefully reading the dataset. The data was the divided, with the training set taking up 70% and the remaining 30% was used for testing (148-training and 64-testing). Cross-validation to prevent overfitting and selection bias is performed during the training of the classifier. 10 subsets were created of the training data which were then tested to arrive at suitable values for parameters to aid classification. The trained classifier was then tested and the results are discussed in section 5. The Support Vector Machine and K-Nearest Neighbors algorithm was used for the evaluation process. The next section provides a short description of the machine learning processes in question. The differences and comparison of the two types of vehicles used (Sedan and SUV) to collect data as well as the results obtained from the classifiers are discussed in detail in section 5.

4. MACHINE LEARNING FOR DATA CLASSIFICATION

Machine Learning is a branch of artificial intelligence, its main objective being “automated generation of knowledge for its incorporation in expert systems”. There are various types of machine learning algorithms applied in civil engineering mainly for data mining (Kotsiantis et al., 2007). In this scenario, machine learning classifiers are used. The first step involves collecting the dataset. From the collected data features are to be extracted which will represent the basis for classification. If the most informative features are not known, then “brute-force” is applied where all the data is extracted hoping a suitable feature can be isolated. Pre-processing is required when dataset is collected by ‘brute-force’. Now the features extracted are plotted in a n-dimensional space. For test data, the user labels it as one particular class. Features belonging to two different classes will show significant variations in values. Once all the test data has been labeled, suitable algorithms are applied to make use of this variations and classify the new data into different sets.

Table 1. Sample DataSet of Accelerometer Readings

LABEL	ROUGH	ROUGH	SMOOTH	UNEVEN	UNEVEN	UNEVEN	UNEVEN	ROUGH
ENTROPY_	2.584963	2.584963	2.584963	2.584963	2.584963	2.584963	2.584963	2.584963
INT-	0.344478	0.088752	0.168259	0.373827	0.098869	0.254659	0.297935	0.051602
PEAK2AV	1.569868	1.588974	1.487198	1.602887	1.552879	1.664184	1.322119	1.928951
PEAK2RM	1.569868	1.588974	1.487198	1.602887	1.552879	1.664184	1.322119	1.928951
KURTOSIS	1.836364	2.7439	2.012388	1.542568	2.293438	1.582068	1.421714	3.404973
SKEWNES	0.166632	0.80325	0.140837	0.222874	0.607136	0.38676	-0.27084	1.191003
VARIANCE	0.041358	0.009271	0.015695	0.034547	0.007525	0.019107	0.02414	0.015881
RMS_f	0.447271	0.26166	0.341735	0.293037	0.236165	0.243351	0.409035	0.222087
PEAKDIFF	0.537362	0.278618	0.350778	0.447406	0.239926	0.333432	0.362673	0.374097
PEAK_f	0.702157	0.41577	0.508228	0.469705	0.366736	0.404982	0.540793	0.428395
STD DEV_f	0.203366	0.096286	0.125281	0.185868	0.086746	0.138228	0.15537	0.126021
MEDIAN_f	0.408992	0.236138	0.329155	0.210627	0.216377	0.17812	0.411873	0.16209
MEAN_f	0.406924	0.246455	0.322031	0.238917	0.222493	0.20808	0.383658	0.189969
ENTROPY	0	0.918296	0	0.918296	1	1.459148	0.650022	1.251629
INT-	3.3812	4.61	1.854	6.5257	9.4239	4.7343	0.8209	1.7692
PEAK2AV	-0.057	1.527698	-0.35541	1.07753	1.403319	1.942609	1.441483	2.061546
PEAK2RM	1.992736	1.527698	1.247763	1.716134	1.44435	1.942609	1.464826	2.061546
KURTOSIS	2.507934	2.138024	2.112621	1.79346	1.74366	2.675766	3.59357	3.407949
SKEWNES	-0.82556	0.503086	0.861249	0.030744	0.233281	0.773215	-1.3272	1.263686
VARIANCE	8.448852	10.59255	1.56204	15.68268	42.54308	21.37741	14.53057	11.9549
RMS	4.098537	3.026253	3.511484	3.712472	6.049019	4.222877	3.868378	3.157242
PEAKDIFF	7.9337	9.0068	3.1335	10.3714	17.2256	13.2591	11.2427	10.027
PEAK	-0.2336	4.6232	-1.248	4.0003	8.4887	8.2034	5.5762	6.5088
STD DEV	2.906691	3.25462	1.249816	3.960137	6.522506	4.623571	3.811899	3.457586
MEDIAN	-2.52245	-1.0984	-3.89365	-1.4905	-1.08325	-0.33065	2.413	-1.03325
MEAN	-3.12367	-0.5754	-3.32097	-0.84472	-1.0668	0.13485	1.689833	-0.07588
Z-Axis	0.77515	0.400667	-0.72775	0.335817	0.091017	1.285733	0.378167	1.610083
Y-Axis	10.9447	10.91057	7.724033	10.336	10.42292	12.9371	10.53682	10.76027
X-Axis	-1.15752	-1.29847	1.97215	1.2082	0.020717	0.257817	0.071767	-2.62657

Note: Here the frequency domain features are represented by '_f'

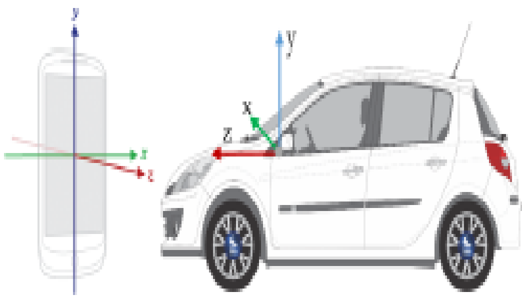

Figure 7. Smartphone and vehicle axes

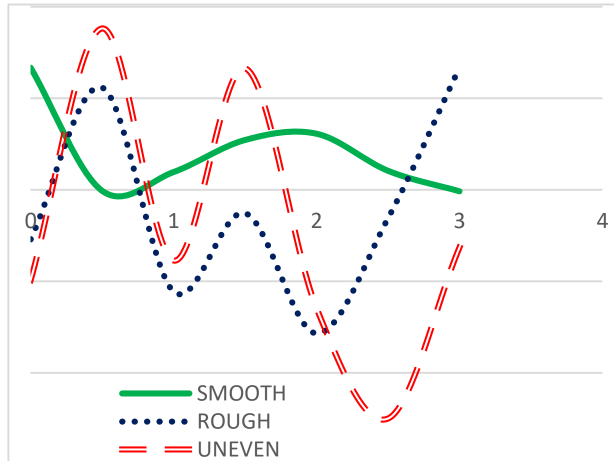

Figure 6. Profile of smooth, rough and uneven pavements.

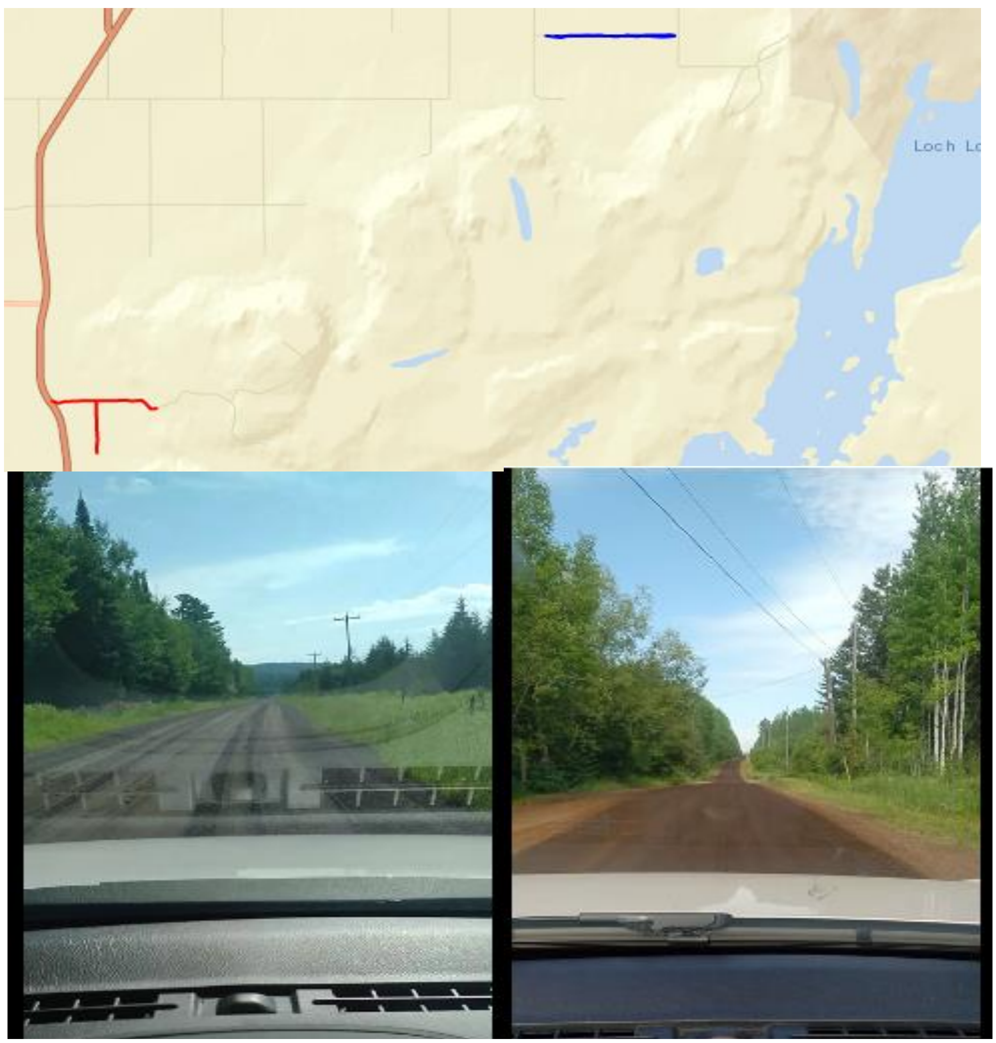

Figure 5. Road map showing Trendaik Road (Blue) and Blake Hall road (Red) and images taken while conducting the test run.

4.1.1. K-NEAREST NEIGHBORS

The KNN is a non-parametric method used for classification and regression (Altman, 1992). The classification rules of KNN are generated by the training data themselves. It is a density based classifier, classifying data based on its k nearest neighbors, working on the principle that data belonging to the same class will have similar features/properties. The data is thrown into a virtual space and the algorithm studies the properties of its k-nearest neighbors, found using Euclidean distances. To report the data point as the label majority of its neighbors possess as seen in Figure 9. The basic logic is fairly simple but it has disadvantages which can be removed with slight modification to its core algorithm (Zhou et al., 2009). The training set should be vast and uniformly distributed to avoid detection problem with the classifier. The main disadvantage being it tends to wrongly classify data belonging to the boundary of a particular label/type as the nearest neighbors may belong to different class.

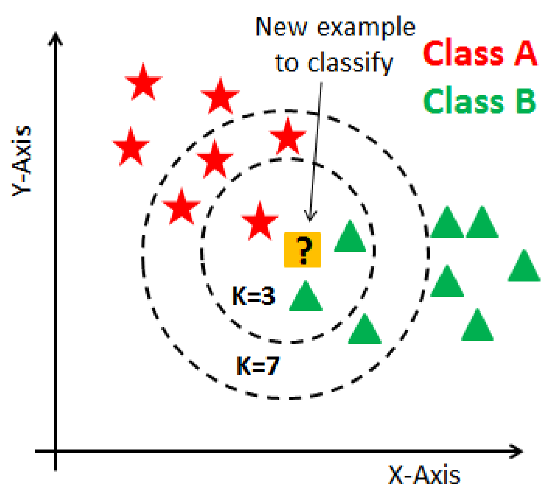

Figure 9. KNN Classification.

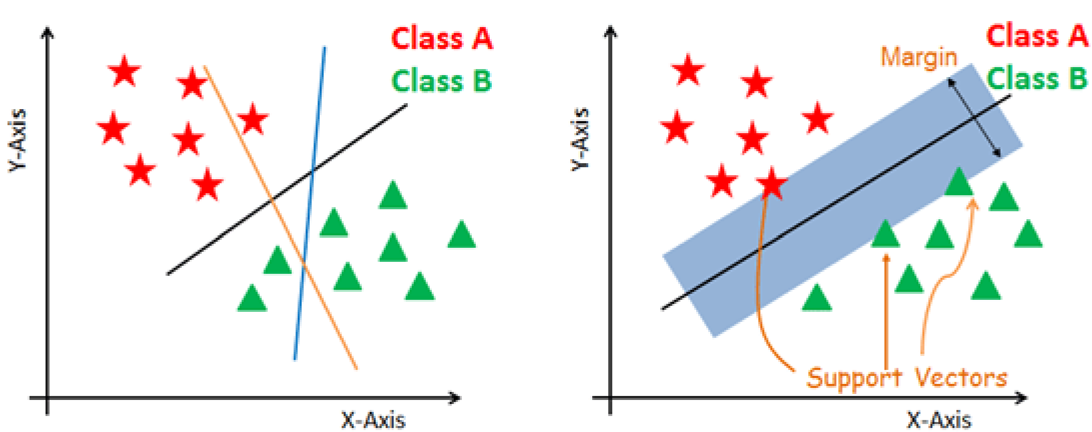

Figure 8. Linear SVM Classification

4.1.2. SUPPORT VECTOR MACHINE

Before the emergence of deep learning techniques, the support vector machines were the most popular algorithm for classification and regression. In SVM, the data is represented in an empty space, mapped so that each category is separated by a clear gap, which is as wide as possible called a separating hyperplane. Now this hyperplane can be in the form of a straight (linear) or a curved line (non-linear). The prediction of the new set is done on the basis on where they are represented in this space which divided into distinct regions belonging to each class separated by the hyperplanes. Figure 8 shows an example of linear classification. The accuracy of SVM depends on a good selection of its hyper-parameters C, γ and kernel parameters. The C parameter eliminates misclassification to some extent, while γ parameter describes the influence of a single training example (Pedregosa et al., 2011).

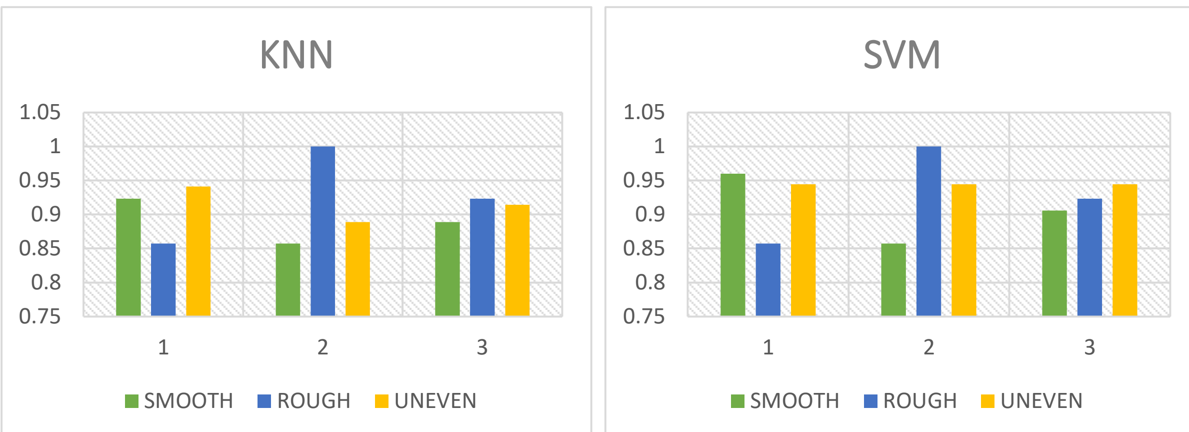

Figure 12. Evaluating criterion for both classifiers (1- Precision, 2- Recall, 3- F1-score)

Figure 11. Owen Drive.

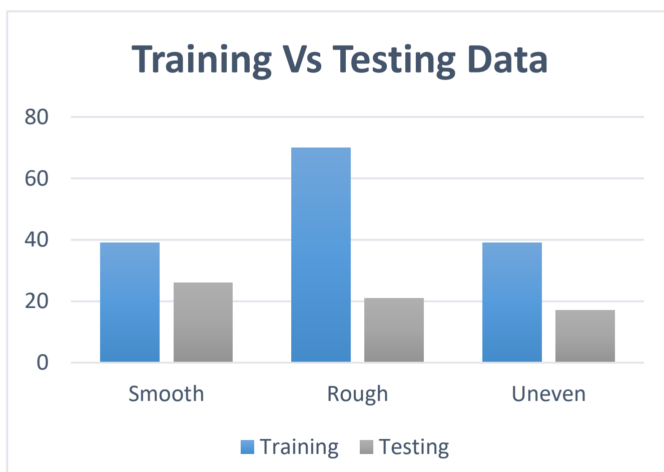

Figure 10. Training Vs Testing Data

5. RESULTS AND DISCUSSION

Previous research has suggested the use and advantages of SVM in pavement classification on paved surfaces (Souza et al., 2018). Even Cabral et al., 2018 reported faster prediction time of 3 seconds for SVM compared to 15 seconds for ResNet, asserting dominance of SVM in real-time classification. To compare its benefits in unpaved road classification, criteria such as precision, recall, accuracy and f1-score were evaluated for KNN and SVM classifiers. As seen in Figure 12 and

Table 2, both classifiers are equally capable of being used in pavement classification, with the accuracy for SVM (92.1875%) being slightly greater than KNN (90.0125%). The classifiers have problem in separating rough and smooth profiles with “smooth” having 4 false negatives and “rough” having 3 false positives for both. The precision, recall and f1-score for smooth profile is 92.31%, 85.71% and 88.89% for KNN and 96%, 85.71% and 90.57% for SVM, while that for uneven profile is 94.12%, 88.89% and 91.43% for KNN and 94.44% for SVM. For rough profile the values are the same for both classifiers, with recall being 100% for rough profile, stating that the algorithm returned the most relevant results in this case. Another explanation can be that since the algorithms had more instances of rough in training data they produce better results in this case (see Figure 10). As is evident from Figure 12 and , higher precision, recall and f1-score for SVM means that it performs better than KNN classifier in this case.

Two different vehicles, a sedan and a SUV were used to collect samples over the same stretch of unpaved road shown in figure 10. On analysis of the video and the accelerometer data it was found that the sedan gave better results compared to SUV. The 0.8 km stretch was traversed twice in both directions giving us a total of 55 instances for both (total – 110 instances). The data was labeled as before. The improved ride quality of the sedan (designed for comfort) compared to the SUV, (designed for rugged terrain) gave the following instances for the same surface, 20 smooth, 20 uneven and 15 rough instances for sedan, while for the SUV 11 smooth, 31 uneven and 14 rough instances. As expected (see Figure 13) the sedan gave less uneven profiles compared to the SUV. The accuracy is predicting these instances was greater in case of SUV (by 8-10%) or both classifiers as the same vehicle was used for collecting the training data (88% for SUV and 80% for sedan).

Table 2. Evaluated Criteria for SVM and KNN classifiers

	KNN			SVM		
	SMOOTH	**ROUGH**	**UNEVEN**	**SMOOTH**	**ROUGH**	**UNEVEN**
TP	24	18	16	24	18	17
TN	34	40	42	35	41	42
FP	2	3	1	1	3	1
FN	4	0	2	4	0	1

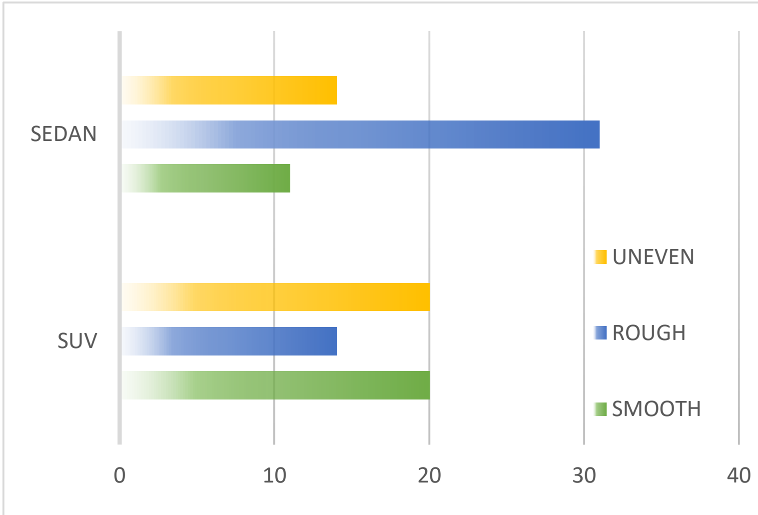

Figure 13. SUV VS SEDAN

6. CONCLUSION

This report presents an effective way to use smartphones for road surface classification, in unpaved roads. The smartphone equipped with an application utilized its various available sensors to record the vibrations while moving on a surface. The data collected was sufficient enough to classify the pavement as rough, smooth and uneven giving over 90% accuracy. The two machine learning techniques compared were K-Nearest Neighbor and Support Vector Machine, with SVM performing better than KNN, emphasizing its superiority. In the field of vehicle telematics other uses of smartphones are GPS navigation, driver risk management, monitoring driver behavior which can also help us in various traffic and safety operations.

It was observed that the road condition was bad especially at locations where two roads merge and on turns, calling for extra care and maintenance in these areas. Also for Blake Hall segment, data was collected before and after maintenance operations. It was seen that the profile for data before maintenance had more uneven segments suggesting that regular maintenance is useful to improve the pavement profile. The use of different vehicles gives us different accelerometer reading. Also the driving speed will alter the data. The purpose of this study was to limit these control parameters and evaluate the efficiency of different algorithms. Further studies should focus on using different vehicles and specific operating speeds to obtain a baseline for the training data to enhance the classification process. Also there lies a need to connect this data/profile observed to already existing measures of pavement roughness such as RQI, Pavement Quality Index and International Roughness Index. The most suitable roughness measure for unpaved roads would be Roughness Defect Score (RDS) and further research needs to be done in this area.

7. REFERENCES

Aksamit, P., Szmechta, M., 2011. Distributed, mobile, social system for road surface defects detection, in: 2011 5th International Symposium on Computational Intelligence and Intelligent Informatics (ISCIII). Presented at the 2011 5th International Symposium on Computational Intelligence and Intelligent Informatics (ISCIII), IEEE, Floriana, Malta, pp. 37–40. https://doi.org/10.1109/ISCIII.2011.6069738

Alessandroni, G., Klopfenstein, L.C., Delpriori, S., Dromedari, M., Luchetti, G., Paolini, B.D., Seraghiti, A., Lattanzi, E., Freschi, V., Carini, A., Bogliolo, A., 2014. SmartRoadSense: Collaborative Road Surface Condition Monitoring 6.

Allouch, A., Koubaa, A., Abbes, T., Ammar, A., 2017. RoadSense: Smartphone Application to Estimate Road Conditions Using Accelerometer and Gyroscope. IEEE Sensors Journal 17, 4231–4238. https://doi.org/10.1109/JSEN.2017.2702739

Altman, N.S., 1992. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. The American Statistician 46, 175–185. https://doi.org/10.1080/00031305.1992.10475879

Cabral, F.S., Pinto, M., Mouzinho, F.A.L.N., Fukai, H., Tamura, S., 2018. An Automatic Survey System for Paved and Unpaved Road Classification and Road Anomaly Detection using Smartphone Sensor, in: 2018 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI). Presented at the 2018 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI), IEEE, Singapore, pp. 65–70. https://doi.org/10.1109/SOLI.2018.8476788

Clyne, T., 2012. Use of accelerated pavement testing to validate Ride Quality Index data. Advances in Pavement Design through Full-scale Accelerated Pavement Testing 387.

Douangphachanh, V., Oneyama, H., 2013. A Study on the Use of Smartphones for Road Roughness Condition Estimation 14.

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. Presented at the Proceedings of the 6th international conference on Mobile systems, applications, and services, ACM, pp. 29–39.

Fwa, T.F., 2005. The Handbook of Highway Engineering. C.R.C Press.

Giroud, J.P., Noiray, L., 1981. Geotextile-reinforced unpaved road design. Journal of Geotechnical and Geoenvironmental Engineering 107(ASCE 16489).

Hanson, T., Cameron, C., Hildebrand, E., 2014. Evaluation of low-cost consumer-level mobile phone technology for measuring international roughness index (IRI) values. Canadian Journal of Civil Engineering 41, 819–827. https://doi.org/10.1139/cjce-2014-0183

Hartgen, D.T., Shufon, J.J., 1983. Windshield Surveys of Highway Condition: A Feasible Input to Pavement Management. Transportation Research Record 9.

Islam, S., Buttlar, W.G., Aldunate, R.G., Vavrik, W.R., 2014. Measurement of Pavement Roughness Using Android-Based Smartphone Application. Transportation Research Record: Journal of the Transportation Research Board 2457, 30–38. https://doi.org/10.3141/2457-04

Kotsiantis, S.B., Zaharakis, I., Pintelas, P., 2007. Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering 160, 3–24.

Lay, M.G., 2009. Handbook of Road Technology. C.R.C Press.

LeClerc, R.V., Marshall, T., 1971. Washington Pavement Rating System: Procedures and Application.

Mednis, A., Strazdins, G., Zviedris, R., Kanonirs, G., Selavo, L., 2011. Real time pothole detection using Android smartphones with accelerometers, in: 2011 International Conference on Distributed Computing in Sensor Systems and Workshops (DCOSS). Presented at the 2011 International Conference on Distributed Computing in Sensor Systems (DCOSS), IEEE, Barcelona, Spain, pp. 1–6. https://doi.org/10.1109/DCOSS.2011.5982206

Mohan, P., Padmanabhan, V.N., Ramjee, R., 2008. Nericell: rich monitoring of road and traffic conditions using mobile smartphones. Presented at the Proceedings of the 6th ACM conference on Embedded network sensor systems, ACM, pp. 323–336.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: Machine learning in Python. Journal of machine learning research 12, 2825–2830.

Pierce, L.M., McGovern, G., Zimmerman, K.A., 2013. Practical Guide for Quality Management of Pavement Condition Data Collection. FHWA.

Radopoulou, S.C., Brilakis, I., 2017. Automated Detection of Multiple Pavement Defects. Journal of Computing in Civil Engineering 31, 04016057. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000623

Sayers, M.W., 1995. On the calculation of international roughness index from longitudinal road profile. Transportation Research Record.

Schnebele, E., Tanyu, B.F., Cervone, G., Waters, N., 2015. Review of remote sensing methodologies for pavement management and assessment. European Transport Research Review 7. https://doi.org/10.1007/s12544-015-0156-6

Selim, A.A., Skorseth, K.O., Muniandy, R., 2003. Long-Lasting Gravel Roads: Case Study from the United States. Transportation Research Record: Journal of the Transportation Research Board 1819, 161–165. https://doi.org/10.3141/1819b-20

Singh, G., Bansal, D., Sofat, S., Aggarwal, N., 2017. Smart patrolling: An efficient road surface monitoring using smartphone sensors and crowdsourcing. Pervasive and Mobile Computing 40, 71–88. https://doi.org/10.1016/j.pmcj.2017.06.002

Souza, V.M.A., Cherman, E.A., Rossi, R.G., Souza, R.A., 2017. Towards Automatic Evaluation of Asphalt Irregularity Using Smartphone's Sensors, in: Adams, N., Tucker, A., Weston, D. (Eds.), Advances in Intelligent Data Analysis XVI. Springer International Publishing, pp. 322–333.

Souza, V.M.A., Giusti, R., Batista, A.J.L., 2018. Asfault: A low-cost system to evaluate pavement conditions in real-time using smartphones and machine learning. Pervasive and Mobile Computing 51, 121–137. https://doi.org/10.1016/j.pmcj.2018.10.008

Strazdins, G., Mednis, A., Kanonirs, G., Zviedris, R., Selavo, L., 2011. Towards Vehicular Sensor Networks with Android Smartphones for Road Surface Monitoring 5.

Tai, Y., Chan, C., Hsu, J.Y., 2010. Automatic Road Anomaly Detection Using Smart Mobile Device 8.

Thompson, R.J., Visser, A.T., 2006. Selection and maintenance of mine haul road wearing course materials. Mining Technology 115, 140–153. https://doi.org/10.1179/174328606X155138

Thompson, R.J., Visser, A.T., 2003. Mine haul road maintenance management systems 10.

WORLD HEALTH ORGANIZATION, 2019. GLOBAL STATUS REPORT ON ROAD SAFETY 2018. WORLD HEALTH ORGANIZATION, S.l.

Yeganeh, S.F., Mahmoudzadeh, A., Azizpour, M.A., Golroo, A., 2019. Validation of smartphone based pavement roughness measures. arXiv:1902.10699 [cs].

Zhou, Y., Li, Y., Xia, S., 2009. An Improved KNN Text Classification Algorithm Based on Clustering. Journal of Computers 4. https://doi.org/10.4304/jcp.4.3.230-237

APPENDIX: MATLAB CODES

1. DEVELOPING FEATURES

```
clear all; close all; clc;
%% Import data from spreadsheet
% Script for importing data from the following spreadsheet:
%
%    Workbook: C:\Users\USER\Desktop\Project\Data\MODIFIED\Blake hall segment
2(MOD).xlsx
%    Worksheet: ACCELEROMETER
%
% To extend the code for use with different selected data or a different
% spreadsheet, generate a function instead of a script.

% Auto-generated by MATLAB on 2019/07/26 23:11:53

%% Import the data
[~, ~, raw0_0] = xlsread('C:\Users\USER\Desktop\Project\Data\MODIFIED\Blake
hall segment 2(MOD).xlsx','ACCELEROMETER','H1:H160');
[~, ~, raw0_1] = xlsread('C:\Users\USER\Desktop\Project\Data\MODIFIED\Blake
hall segment 2(MOD).xlsx','ACCELEROMETER','J1:J160');
raw = [raw0_0,raw0_1];
raw(cellfun(@(x) ~isempty(x) && isnumeric(x) && isnan(x),raw)) = {''};

%% Replace non-numeric cells with NaN
R = cellfun(@(x) ~isnumeric(x) && ~islogical(x),raw); % Find non-numeric
cells
raw(R) = {NaN}; % Replace non-numeric cells

%% Create output variable
data = reshape([raw{:}],size(raw));

%% Create table
Blakehallsegment2MODS1 = table;

%% Allocate imported array to column variable names
ACCELERATIONY = data(:,1);
Time = data(:,2);

%% Clear temporary variables
clearvars data raw raw0_0 raw0_1 R;

%% FFT to frequency domain

Fs=2;
T=1/Fs;
ACCELERATIONY(isnan(ACCELERATIONY))=[];
Time(isnan(Time))=[];
L=length(ACCELERATIONY);
FFTY=fft(ACCELERATIONY)/L;
FFTY_mag=abs(FFTY);
fv=linspace(0,Fs,fix(L));
iv=1:length(fv);
```

```
fileName = 'FDBLAKEHALLS2.xlsx';
sheet=1;
xlswrite(fileName, FFTY_mag, sheet, 'A2');
xlswrite(fileName, fv', sheet, 'B2');

%% Plotting graphs

figure (1);
plot(fv, FFTY_mag(iv));
xlabel('Frequency(Hz)');
ylabel('Amplitude');
title('Frequency Domain');
figure(2);
plot(Time/1000, ACCELERATIONY);
xlabel('Time (s)');
ylabel('Y-Axis Accelerometer data');
title ('Time Domain');

%% VARIABLES

MEAN=zeros(100,1); MEDIAN=zeros(100,1); STDEVIATION = zeros(100,1);
PEAK =zeros(100,1);PEAKDIFF =zeros(100,1);
ROOTMEANSQUARE=zeros(100,1); VARIANCE=zeros(100,1); SKEWNESS=zeros(100,1);
KURTOSIS =zeros(100,1); PEAK2RMS =zeros(100,1);
PEAK2AVG =zeros(100,1); inquarange=zeros(100,1);
ENTROPY =zeros(100,1); MEAN_f =zeros(100,1);
MEDIAN_f =zeros(100,1); STDEVIATION_f =zeros(100,1);
PEAK_f =zeros(100,1); PEAKDIFF_f =zeros(100,1);
ROOTMEANSQUARE_f =zeros(100,1); VARIANCE_f =zeros(100,1);
SKEWNESS_f =zeros(100,1); KURTOSIS_f =zeros(100,1);
PEAK2RMS_f =zeros(100,1); PEAK2AVG_f =zeros(100,1);
inquarange_f =zeros(100,1); ENTROPY_f =zeros(100,1);

%% Temporal Features
j=1;b=1;
ACCELERATIONY(isnan(ACCELERATIONY))=[];AMPLITUDE=FFTY_mag;
for i=1:length(ACCELERATIONY)
    b=mod(i,6);
        if(b==0)
        MEAN(j)=mean(ACCELERATIONY(i-5:i));
        MEDIAN(j)=median(ACCELERATIONY(i-5:i));
        STDEVIATION(j)=std(ACCELERATIONY(i-5:i));
        PEAK(j)=max(ACCELERATIONY(i-5:i));
        PEAKDIFF(j)=PEAK(j)-min(ACCELERATIONY(i-5:i));
        ROOTMEANSQUARE(j)=rms(ACCELERATIONY(i-5:i));
        VARIANCE(j)=var(ACCELERATIONY(i-5:i));
        SKEWNESS(j)=skewness(ACCELERATIONY(i-5:i));
        KURTOSIS(j)=kurtosis(ACCELERATIONY(i-5:i));
        PEAK2RMS(j)=peak2rms(ACCELERATIONY(i-5:i));
        PEAK2AVG(j)= PEAK(j)/ROOTMEANSQUARE(j);
        inquarange(j)=iqr(ACCELERATIONY(i-5:i));
        ENTROPY(j)=entropy(ACCELERATIONY(i-5:i));
        MEAN_f(j)=mean(AMPLITUDE(i-5:i));
        MEDIAN_f(j)=median(AMPLITUDE(i-5:i));
        STDEVIATION_f(j)=std(AMPLITUDE(i-5:i));
        PEAK_f(j)=max(AMPLITUDE(i-5:i));
```

```
        PEAKDIFF_f(j)=PEAK_f(j)-min(AMPLITUDE(i-5:i));
        ROOTMEANSQUARE_f(j)=rms(AMPLITUDE(i-5:i));
        VARIANCE_f(j)=var(AMPLITUDE(i-5:i));
        SKEWNESS_f(j)=skewness(AMPLITUDE(i-5:i));
        KURTOSIS_f(j)=kurtosis(AMPLITUDE(i-5:i));
        PEAK2RMS_f(j)=peak2rms(AMPLITUDE(i-5:i));
        PEAK2AVG_f(j)= PEAK_f(j)/ROOTMEANSQUARE_f(j);
        inquarange_f(j)=iqr(AMPLITUDE(i-5:i));
        ENTROPY_f(j)=entropy(AMPLITUDE(i-5:i));
        j=j+1;
        end
end
if(b>0)
n=(j-1)*6;
MEAN_f(j)=mean(AMPLITUDE((n+1):(n+b)));
MEDIAN_f(j)=median(AMPLITUDE((n+1):(n+b)));
STDEVIATION_f(j)=std(AMPLITUDE((n+1):(n+b)));
PEAK_f(j)=max(AMPLITUDE((n+1):(n+b)));
PEAKDIFF_f(j)=PEAK_f(j)-min(AMPLITUDE((n+1):(n+b)));
ROOTMEANSQUARE_f(j)=rms(AMPLITUDE((n+1):(n+b)));
VARIANCE_f(j)=var(AMPLITUDE((n+1):(n+b)));
SKEWNESS_f(j)=skewness(AMPLITUDE((n+1):(n+b)));
KURTOSIS_f(j)=kurtosis(AMPLITUDE((n+1):(n+b)));
PEAK2RMS_f(j)=peak2rms(AMPLITUDE((n+1):(n+b)));
PEAK2AVG_f(j)= PEAK_f(j)/ROOTMEANSQUARE_f(j);
inquarange_f(j)=iqr(AMPLITUDE((n+1):(n+b)));
ENTROPY_f(j)=entropy(AMPLITUDE((n+1):(n+b)));
MEAN(j)=mean(ACCELERATIONY((n+1):(n+b)));
MEDIAN(j)=median(ACCELERATIONY((n+1):(n+b)));
STDEVIATION(j)=std(ACCELERATIONY((n+1):(n+b)));
PEAK(j)=max(ACCELERATIONY((n+1):(n+b)));
PEAKDIFF(j)=PEAK(j)-min(ACCELERATIONY((n+1):(n+b)));
ROOTMEANSQUARE(j)=rms(ACCELERATIONY((n+1):(n+b)));
VARIANCE(j)=var(ACCELERATIONY((n+1):(n+b)));
SKEWNESS(j)=skewness(ACCELERATIONY((n+1):(n+b)));
KURTOSIS(j)=kurtosis(ACCELERATIONY((n+1):(n+b)));
PEAK2RMS(j)=peak2rms(ACCELERATIONY((n+1):(n+b)));
PEAK2AVG(j)= PEAK(j)/ROOTMEANSQUARE(j);
inquarange(j)=iqr(ACCELERATIONY((n+1):(n+b)));
ENTROPY(j)=entropy(ACCELERATIONY((n+1):(n+b)));
end
%% Writing to excel file

fileName='BLAKEHALLS2_accelero_FEATURES.xlsx';
sheet = 1;
xlswrite(fileName, MEAN, sheet,'C2');
xlswrite(fileName, MEDIAN, sheet,'D2');
xlswrite(fileName, STDEVIATION, sheet,'E2');
xlswrite(fileName, PEAK, sheet,'F2');
xlswrite(fileName, PEAKDIFF, sheet,'G2');
xlswrite(fileName, ROOTMEANSQUARE, sheet,'H2');
xlswrite(fileName, VARIANCE, sheet,'I2');
xlswrite(fileName, SKEWNESS, sheet,'J2');
xlswrite(fileName, KURTOSIS, sheet,'K2');
xlswrite(fileName, PEAK2RMS, sheet,'L2');
xlswrite(fileName, PEAK2AVG, sheet,'M2');
xlswrite(fileName, inquarange, sheet,'N2');
```

```
xlswrite(fileName, ENTROPY,sheet,'O2');
fileName='FDBLAKEHALLS2.xlsx';
sheet = 2;
xlswrite(fileName, MEAN_f, sheet,'C2');
xlswrite(fileName, MEDIAN_f, sheet,'D2');
xlswrite(fileName, STDEVIATION_f, sheet,'E2');
xlswrite(fileName, PEAK_f, sheet,'F2');
xlswrite(fileName, PEAKDIFF_f, sheet,'G2');
xlswrite(fileName, ROOTMEANSQUARE_f, sheet,'H2');
xlswrite(fileName, VARIANCE_f, sheet,'I2');
xlswrite(fileName, SKEWNESS_f, sheet,'J2');
xlswrite(fileName, KURTOSIS_f, sheet,'K2');
xlswrite(fileName, PEAK2RMS_f, sheet,'L2');
xlswrite(fileName, PEAK2AVG_f, sheet,'M2');
xlswrite(fileName, inquarange_f, sheet,'N2');
xlswrite(fileName, ENTROPY_f, sheet,'O2');
```

2. CREATING DATASET

```
clear; close; clc;
load ANNDATA.mat;
%% Define Testing and Sample Data
k=randperm(212);ri=1;ui=1;si=1;
X_Samp=table2array(ANNDATA1(k(1:148),1:29));
label=grp2idx(LABEL);
X_Test=table2array(ANNDATA1(k(149:end),1:29));
Y_samp=grp2idx(LABEL(k(1:148)));
Y_test=grp2idx(LABEL(k(149:end)));
for i=1:length(Y_samp)
    if (Y_samp(i)==1)
        ri=ri+1;
    else if (Y_samp(i)==2)
            si=si+1;
        else
            ui=ui+1;
        end
    end
end
```

3. SVM

```
clear; close; clc;
load 'ML_DAT.mat';

%% CV Partition for cross validation
c=cvpartition(Y_samp, 'k', 5);

%% Train Data

MD1=fitcecoc(X_Samp, Y_samp, 'Coding', 'ordinal', 'ClassNames',[1,2,3],....
    'CVPartition', c,'Learners','svm'),

%% Selecting Best Parameters

MD1=fitcecoc(X_Samp, Y_samp, 'OptimizeHyperparameters', 'auto',.....
    'HyperparameterOptimizationOptions',struct('AcquisitionFunctionName',....
```

```
    'expected-improvement-plus')),

%% Predicting outputs

TNU=0;TPS=0;TPR=0;TPU=0;FNR=0;FNS=0;FNU=0;TNS=0;TNR=0;TNP=0;FPS=0;FPR=0;FPU=0
;
label = predict(MD1,X_Test);
for i=1:length(label)
    if (Y_test(i)==1)
        if(label(i)==Y_test(i))
            TPS=TPS+1;
            TNR=TNR+1;
            TNU=TNU+1;
        else if (label(i)==2)
                FPR=FPR+1;
                FNS=FNS+1;
            else
                FPU=FPU+1;
                FNS=FNS+1;
            end
        end
    end
    if (Y_test(i)==2)
        if(label(i)==Y_test(i))
            TPR=TPR+1;
            TNS=TNS+1;
            TNU=TNU+1;
        else if (label(i)==1)
                FPS=FPS+1;
                FNR=FNR+1;
            else
                FPU=FPU+1;
                FNR=FNR+1;
            end
        end
    end
        if (Y_test(i)==3)
            if(label(i)==Y_test(i))
                TPU=TPU+1;
                TNS=TNS+1;
                TNR=TNR+1;
            else if (label(i)==2)
                FPR=FPR+1;
                FNU=FNU+1;
            else
                FPS=FPS+1;
                FNU=FNU+1;
                end
            end
        end
end
accuracy= (sum(label==Y_test)/length(Y_test))*100,
PrecisionSmooth=TPS/(TPS+FPS),
PrecisionRough=TPR/(TPR+FPR),
PrecisionUneven=TPU/(TPU+FPU),
RecallSmooth=TPS/(TPS+FNS),
```

```
RecallRough=TPR/(TPR+FNR),
RecallUneven=TPU/(TPU+FNU),
FScoreSmooth=2/((1/RecallSmooth)+(1/PrecisionSmooth)),
FScoreRough=2/((1/RecallRough)+(1/PrecisionRough)),
FScoreUneven=2/((1/RecallUneven)+(1/PrecisionUneven)),

%% Plotting Hyperplane

gscatter(X_Test(:,7),X_Test(:,8),label,'rbg');
legend('Smooth','Rough','Uneven');

%% SEDAN&SUV

load 'OwenDrive.mat'
sedan=predict(MD1,SedanData);
suv=predict(MD1,SUVData);
SUVAcc= (sum(sedan==SedanLabel)/length(SedanLabel))*100;
SEDacc= (sum(suv==SUVlabel)/length(SUVlabel))*100;
```

4. KNN

```
clear; close; clc;
load ML_DAT.mat;

%% CV Partition for cross validation
c=cvpartition(Y_samp, 'k', 5);

%% Train Data

MD1=fitcecoc(X_Samp, Y_samp, 'Coding', 'ordinal', 'ClassNames',[1,2,3],....
    'CVPartition', c,'Learners','knn'),

%% Selecting Best Parameters

MD1=fitcecoc(X_Samp, Y_samp, 'OptimizeHyperparameters', 'auto',.....
    'HyperparameterOptimizationOptions',struct('AcquisitionFunctionName',....
    'expected-improvement-plus')),

%% Predicting outputs

TNU=0;TPS=0;TPR=0;TPU=0;FNR=0;FNS=0;FNU=0;TNS=0;TNR=0;TNP=0;FPS=0;FPR=0;FPU=0
;
label = predict(MD1,X_Test);
for i=1:length(label)
    if (Y_test(i)==1)
        if(label(i)==Y_test(i))
            TPS=TPS+1;
            TNR=TNR+1;
            TNU=TNU+1;
        else if (label(i)==2)
                FPR=FPR+1;
                FNS=FNS+1;
            else
```

```
                FPU=FPU+1;
                FNS=FNS+1;
            end
        end
    end
    if (Y_test(i)==2)
        if(label(i)==Y_test(i))
            TPR=TPR+1;
            TNS=TNS+1;
            TNU=TNU+1;
        else if (label(i)==1)
                FPS=FPS+1;
                FNR=FNR+1;
            else
                FPU=FPU+1;
                FNR=FNR+1;
            end
        end
    end
        if (Y_test(i)==3)
            if(label(i)==Y_test(i))
                TPU=TPU+1;
                TNS=TNS+1;
                TNR=TNR+1;
            else if (label(i)==2)
                FPR=FPR+1;
                FNU=FNU+1;
            else
                FPS=FPS+1;
                FNU=FNU+1;
                end
            end
        end
end
accuracy= (sum(label==Y_test)/length(Y_test))*100,
PrecisionSmooth=TPS/(TPS+FPS),
PrecisionRough=TPR/(TPR+FPR),
PrecisionUneven=TPU/(TPU+FPU),
RecallSmooth=TPS/(TPS+FNS),
RecallRough=TPR/(TPR+FNR),
RecallUneven=TPU/(TPU+FNU),
FScoreSmooth=2/((1/RecallSmooth)+(1/PrecisionSmooth)),
FScoreRough=2/((1/RecallRough)+(1/PrecisionRough)),
FScoreUneven=2/((1/RecallUneven)+(1/PrecisionUneven)),
%% Plot

gscatter(X_Test(:,7),X_Test(:,8),label,'rbg');
legend('Smooth','Rough','Uneven');

%% SEDAN&SUV

load 'OwenDrive.mat'
sedan=predict(MD1,SedanData);
suv=predict(MD1,SUVData);
SUVAcc= (sum(sedan==SedanLabel)/length(SedanLabel))*100;
SEDacc= (sum(suv==SUVlabel)/length(SUVlabel))*100;
```

5. PLOTTING THE COORDINATES ON MAP

```
webmap;
load Mapping.mat;
wmline(TSegment1Lat, TSegment1Long,'Color', 'blue','LineWidth', 2);
wmline(TSegment2Lat, TSegment2Long,'Color', 'blue','LineWidth', 2);
wmline(TReturnLat, TReturnLong, 'Color', 'blue','LineWidth', 2);
wmline(BHSegment1Lat, BHSegment1Long,'Color', 'red','LineWidth', 2);
wmline(BHSegment2Lat, BHSegment2Long,'Color', 'red','LineWidth', 2);
wmline(OwenDrLat, OwenDrLong,'Color', 'yellow','LineWidth', 2);
wmline(OwenDReturnLat, OwenDReturnLong,'Color', 'yellow','LineWidth', 2);
```