



Measuring Road Roughness through Crowdsourcing while Minimizing the Conditional Effects

Y. T. Gamage¹ · T. A. I. Thotawaththa¹ · A. Wijayasiri¹

Received: 6 January 2022 / Revised: 7 May 2022 / Accepted: 8 June 2022 / Published online: 28 June 2022
© The Author(s), under exclusive licence to Intelligent Transportation Systems Japan 2022

Abstract

A well-maintained road network is a crucial factor for sustainable urban development. Over the past few years, researchers have proposed smartphone-based crowdsourced applications as a low-cost effective solution to acquire frequent road surface quality updates. One of the main limitations faced by these applications is that the collected values exhibit significant variations over the conditions under which the road data was collected. This study is an attempt to develop a road roughness monitoring platform using passenger cars that can produce accurate results while reducing the effect of these conditions such as the car type, smartphone model, or its placement. The developed system consists of several features including automatic journey detection, freedom to use any smartphone in any position with or without an active internet connection when collecting data, converging values collected from different sources, and visualizing them in a virtual map. A set of field tests were carried out to evaluate the proposed system based on the road condition, passenger car type, smartphone model, and smartphone placement inside the vehicle. The results show that the proposed solution is effective in predicting accurate values after reducing the effect of these varying factors.

Keywords Road roughness · International roughness index · Signal processing · Machine learning · Crowdsourcing

1 Introduction

Infrastructure maintenance is paramount for sustainable urbanization. With the population residing in urban areas doubling every year [15], the consequent degradation of infrastructure is inevitable. If urbanization happens unplanned it will exert negative impacts on a country's economic growth, social welfare, and overall quality of life [8]. The smart city concept comes into play as an approach to find solutions to problems arising from rapid urbanization, with sustainability in mind. In a broad definition, a smart city is about creating new ways of incorporating collaborations from individuals for the betterment of city governance with the use of information

and communication technology (ICT) [30]. The innovations and expansions of ICT expedite the growth of smart cities.

Among the other types of infrastructure, the road network plays a vital role in sustainable urban development. Technical defects and poor road conditions result in poor ride quality and additional costs due to increased fuel consumption, vehicle damages, and additional travel time. As fuel consumption is directly related to the vehicle's CO₂ emission [18], it also triggers environmental pollution. Therefore road conditions should be frequently monitored to ensure they comply with the correct standards. The traditional methods of road quality monitoring such as the use of dipstick profilers, bump integrators and visual inspection are labour-intensive, expensive, and require expert supervision. Developing countries with emerging economies struggle to find such a workforce or precise technologies to maintain their rapidly expanding road system. In Sri Lanka as well the road sector has been expanding over the years resulting in becoming the country with the highest road density in South Asia [27]. However, around 25% of this road network suffers from negligence and lack of maintenance especially in rural areas [3].

✉ Y. T. Gamage
yogyagamage.17@cse.mrt.ac.lk
T. A. I. Thotawaththa
achinthaisuru.17@cse.mrt.ac.lk
A. Wijayasiri
adeeshaw@cse.mrt.ac.lk

¹ University of Moratuwa, Katubedda 10400, Sri Lanka

Previous research has revealed that smartphone-based solutions can be used as low-cost and robust techniques to monitor road quality [6, 16, 17]. These applications use accelerometers and magnetometers embedded in modern smartphones to collect road data. As the number of smartphone users is rapidly increasing every year, these methods create a great opportunity to frequently collect and update road quality status by crowdsourcing. However, crowdsourcing can incur errors and inconsistencies in calculated values as the collected data can vary based on the vehicle type, smartphone model, or other user-specific habits of using mobile phones. This research aims to provide a solution to minimize these errors and efficiently employ crowdsourcing techniques.

The main contributions of this paper can be identified as follows.

- A novel solution to measure road surface quality which is less susceptible to various factors that deviate the predicted roughness level from the actual level, and expects minimum requirements from the user for data collection is presented.
- A segmentation algorithm is presented to segmentize any road using vehicle journey routes without having predefined segments, and to identify repeated segments in multiple journeys.
- An updating algorithm is introduced to calculate a balanced roughness value from the acquired data on different journeys while penalizing possible errors. Moreover, a dashboard with a virtual map of roads is developed using Google Map APIs to display the calculated values.

2 Related Work

A variety of methods to assess road surface quality have been explored in the literature. In recent years, a multitude of research studies has been conducted under smartphone-based sensing approaches. Roadroid [21] can be identified as one of the earliest software developed to monitor road conditions using smartphone sensors. According to the authors, smartphone-based monitoring falls into class 4, which is subjective rating and uncalibrated measures, and would be a powerful way of data collection even though it will not provide precise results compared to the expensive precision profiles. They have experimented on the influence of the different vehicle types and road types, and have developed a model to make calculations less susceptible. Their two options of roughness calculation, eIRI (estimated IRI) and cIRI (calculated IRI) were based on peaks and root mean square analysis of vibrations and quarter-car-simulation.

Several research studies have investigated the use of machine learning techniques to predict IRI. Asfaut, a mobile application, and a crowdsourced platform to monitor

road conditions in real-time using machine learning algorithms have been developed under the research study referred to in [36]. They have considered the evaluation of road condition as a 5 class classification problem (good, average, fair, poor, and occurrences of obstacles). The SVM (Support Vector Machines) classifier was used for the automated evaluation of road quality and the average accuracy was 92%. However, the data collection was only available to the developers or authorized experts. RoadSense [7] is an Android application that makes use of both accelerometer and gyroscope sensors. The authors have experimented with three machine learning algorithms, C4.5 decision tree, SVM, and Naive Bayes.

In [34] the authors have provided a comprehensive review of the existing literature on smartphone sensor-based approaches for road quality monitoring. The conclusions drawn from the review include the necessity of developing a fully automated application without the involvement of user interaction, providing adaptability to variations observed under different vehicle types and road conditions, and applying effective crowdsourcing techniques.

Experimental studies to identify and evaluate the factors that affect the road quality measurements of smartphones have been carried out in [5, 29, 40]. They have investigated the influence of speed, the mobile phone model, its placement, vehicle type and driving style. The results of the experiments suggest that without proper calibration the obtained values can vary significantly, however by applying suitable techniques and analyzing data from all sources together, the difference can be minimized.

Crowdsourced techniques help to overcome the problems of insufficient data coverage and can provide more frequent updates, presumably improving the accuracy of the observations. In [24] the authors have proposed an optimized crowdsensing solution to assess road quality. Reorientation mechanisms to align the smartphone accelerometer with the vehicle axes were applied through the use of Euler Angles. An IRI proxy value was calculated using the Root Mean Square of the z-axis acceleration, the number of values collected in the 50m segment, and vehicle speed. Data from multiple users were integrated by simply merging it to the nearest existing point using a distance-based weighting scheme.

A road condition monitoring system called RCT (Road Condition Tool) is proposed in [38] and its performance was analyzed in [37]. The objective of the application was to support supply chain links in identifying appropriate routes. A set of parameters including the vehicle type and weight along with acceleration and GPS data were collected. The repeatability of the calculated road pavement condition assessment index was analyzed in terms of vehicle running speed and vehicle category. Except for the category of heavy goods vehicles, the other vehicle categories and the vehicle speed had shown a low correlation with the index.

The main drawbacks of these previously proposed solutions are, requiring user interactions when collecting vehicle journey data and not handling smartphone sensor noises and errors efficiently to make the solutions properly crowdsourced. Moreover, they have not proposed proper techniques to recognize repeated road segments or to merge collected values from different sources accordingly. The majority of the previous solutions also lack validation of the solution across varying conditions. Table 1 presents a summary of the previous literature on smartphone-based road condition monitoring.

3 Methodology

3.1 Overview of the Data Collecting Mobile Application

An Android application “iRoads X” which is an updated version of “iRoads” [4] is used to collect GPS and acceleration data. The collected data; which are anonymised to protect the privacy of the user; are filtered, preprocessed, stored, and transferred to a database. Figure 1 shows two user interfaces of the iRoads X application. The rest of this section describes the techniques used in the iRoads X application for data collection, preprocessing, and storing.

3.2 Speed Calculation

iRoads X uses GPS to calculate the speed of the vehicle. It incorporates the haversine formula [14] given in Eq. 1, to get the distance between two GPS coordinates, and then that distance is divided by the time taken to travel the distance.

$$d = 2r \arcsin \left(\sqrt{\sin^2 \left(\frac{\phi_2 - \phi_1}{2} \right) + \cos(\phi_1) \cos(\phi_2) \sin^2 \left(\frac{\lambda_2 - \lambda_1}{2} \right)} \right) \tag{1}$$

- r is the radius of the Earth(i.e. 6371 km)
- ϕ_1, ϕ_2 are the latitudes of point 1 and point 2 (in radians)
- λ_1, λ_2 are the longitudes of point 1 and point 2 (in radians)

3.3 Noise Filtering

When collecting smartphone sensor data by crowdsourcing, the acquired signals may contain various types of noises. The SMA (Simple Moving Average) filter is used by iRoads X to decrease noises in the 3-axis accelerometer, such as hand movements and vehicle maneuvers. SMA is an arithmetic moving average that is calculated by taking the average of a data series. It can be mathematically represented by Eq. 2. In this equation A_i is the data point in the i th period, and n is the total number of periods.

$$\frac{A_1 + A_2 + \dots + A_n}{n} \tag{2}$$

Then the filtered data is put through a reorientation process which is explained in the following section. To remove constant noise such as engine vibrations, the system maintains a queue for the average filters of each axis of the accelerometer. Subsequently, another queue is maintained to keep reoriented sensor data collected when the vehicle is stable. Then the mean value of this queue is reduced from the mean value obtained from the previous queue to obtain the average filtered constant noise from the accelerometer readings.

3.4 Re-Orientation

iRoads X uses the 3-axis accelerometer as the main sensor to collect road profile data. A reorientation mechanism is required as the users can keep the mobile phone in any position while collecting the accelerometer data. For this reorientation process, configurable two methods are used in the iRoads X application. One method is by calculating Euler angles [31]. Equations 3 to 13 represent the mathematical functions involved in this method. $x, y,$ and z are the initial acceleration values along $x, y,$ and z axes and $x_{reoriented}, y_{reoriented},$ and $z_{reoriented}$ are the accelerations after reorienting.

$$\theta = \arccos \left(\frac{y}{9.800} \right) \tag{3}$$

$$\phi = \arctan \left(\frac{z}{x} \right) \tag{4}$$

$$x_\phi = x \times \cos(\phi) - z \times \sin(\phi) \tag{5}$$

$$y_\phi = y \tag{6}$$

$$z_\phi = x \times \sin(\phi) + z \times \cos(\phi) \tag{7}$$

$$x_\theta = x_\phi \times \cos(\theta) + y_\phi \times \sin(\theta) \tag{8}$$

$$z_\theta = z_\phi \tag{9}$$

$$\alpha = \arctan \left(\frac{x_\phi}{z_\phi} \right) \tag{10}$$

$$x_{reoriented} = x_\theta \times \cos(\alpha) - z_\theta \times \sin(\alpha) \tag{11}$$

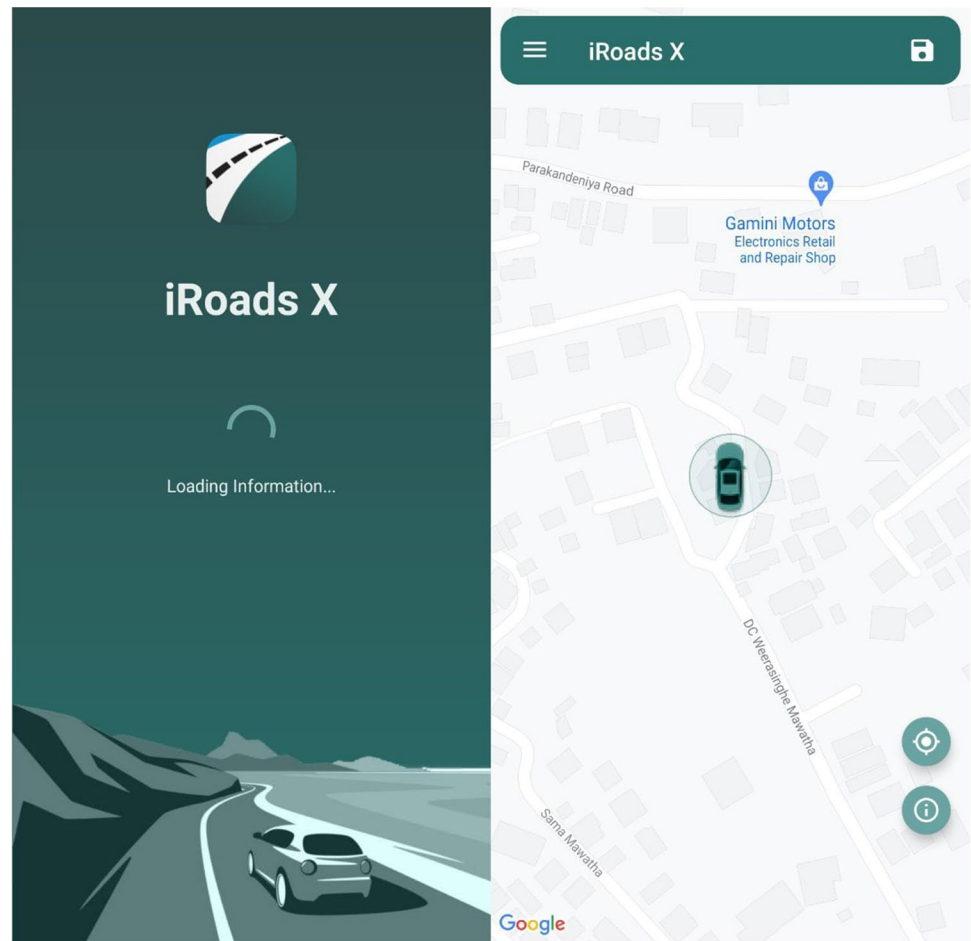
$$y_{reoriented} = -x_\phi \times \sin(\theta) + y_\phi \cos(\theta) \tag{12}$$

$$z_{reoriented} = x_\theta \times \sin(\alpha) + z_\theta \times \cos(\alpha) \tag{13}$$

Table 1 A comparison of previous road condition monitoring smartphone applications

Reference	Techniques	Roughness measure	Crowdsourced?	Field-tests	Segmentation approach	Updating method
[33]	SVM	Anomaly detection - railroad crossings, potholes, bumps, etc.	No	Using Nokia N95 8GB placed on the windshield rack. Vehicle - not mentioned	Segments selected based on distance travelled during a specified time period	None
[21]	Regression based approach	Measure eIRI and cIRI, then categorize under 4 classes, Good, Ok, NotOK, Poor	No - data collection is expected to be done by officers	Using Samsung Galaxy Tab GT P1000 on different vehicles (small car to the large 4WD jeep)	IRI values were saved in a point-based manner and for a given segment size the average was given using the saved points	Take the average
[20]	Using ProVal Software	Predict IRI	No	On three sites with different roughness levels using a profiler van and a passenger car (Mazda3)	Predefined 0.1-mile segments	None
[36]	SVM	5 class classification problems (good, average, fair, poor, and occurrences of obstacles)	Data collection feature available only for authorized experts	Using Samsung Galaxy S7, S8, and Motorola G5 and a medium-sized hatchback car (Hyundai i30).	Segments selected based on distance travelled during a specified time period (5s)	None
[7]	Three machine learning algorithms - C4.5 decision tree, SVM, and Naive Bayes	Classified as Smooth or Potholed	No	Using Samsung galaxy alpha. Vehicle - not mentioned	Segments selected based on distance travelled during a specified time period	None
[25]	Threshold-based	Classified as good, normal, bad, terrible	Yes	Vehicle or mobile phone types not mentioned	By drawing rectangular polygons around each segment	Take the average
[22]	Threshold-based, fuzzy logic approach	Classified as Smooth, Some discomfort, Moderate discomfort, Major discomfort, Very rough	Yes	Five smartphones of popular brands on ten vehicles which were mainly two-wheelers. (Specific types not mentioned)	Defined 100m segments on predefined routes	Average using the fuzzy system
[24]	Equation-threshold based	IRI proxy value and number of transient events	Yes	iPhone 6, iPhone 7, and Moto X Pure fixed in the cabin of a 2009 Toyota Corolla	Routes segmented for 50m sections (road changes are not considered)	Merge new point values to the nearest existing point using a distance-based weighting scheme
[38]	Equation - threshold-based	Classified under 4 classes, Good, Satisfactory, Unsatisfactory, Poor	Yes - for freight vehicles	On a passenger car, a delivery van, and a heavy goods vehicle	Fixed 10m sections on predefined routes	Take the arithmetic mean of 30 days or 1 day or 1 week (under conditions)

Fig. 1 User interfaces of the iRoads X application



The other method is by using the magnetic vector and the gravity vector which can be found using the magnetometer and the accelerometer in the smartphone [9]. In this method, the reorientation can be performed without waiting for braking events. First, the accelerometer sensor data is transformed from the smartphone coordinate system to the geometric coordinate system and then transformed to the vehicle's coordinate system. Figure 2 demonstrates this reorientation process.

3.5 Identifying a Journey Without User Input

Another feature that iRoads X utilizes to support crowdsourcing is automatically identifying the time to collect data; i.e., collecting data only when the user is travelling in a vehicle. For this, iRoads X uses a third-party Software Development Kit (SDK) called pathsense [2] to identify the start and the end of a vehicle journey. This feature can minimize the user interactions required for the data collection process, which makes iRoads X more suitable and user-friendly for crowdsourcing.

3.6 Database

As the local database to store the collected data, iRoads X uses Couchbase lite [1]. The stored data is communicated with the database server through a sync gateway [32]. This provides an additional advantage that an active internet connection is not required to collect the data, the users can connect later and sync the data in the local database with the database server. The overall architecture of the application is presented in Fig. 3.

3.7 Segmentation of Roads

A segmentation algorithm is implemented to segmentize any road, and to identify the segments which are repeatedly travelled by vehicles using GPS coordinates. The roads do not need to have predefined segments, they will be automatically segmented using vehicle routes. This section presents the implementation methodology of the segmentation algorithm.

As the initial step, the collected journey data sorted by the timestamp are retrieved from the couchbase server. Then

Fig. 2 Reorientation of the mobile phone

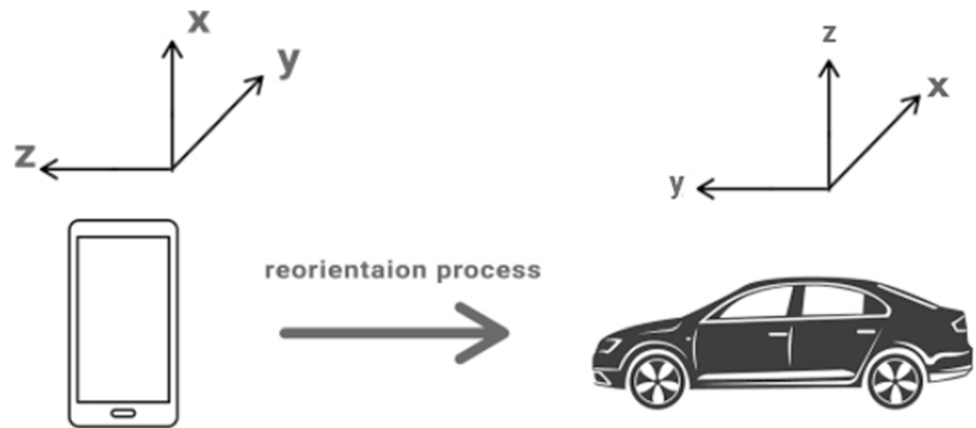
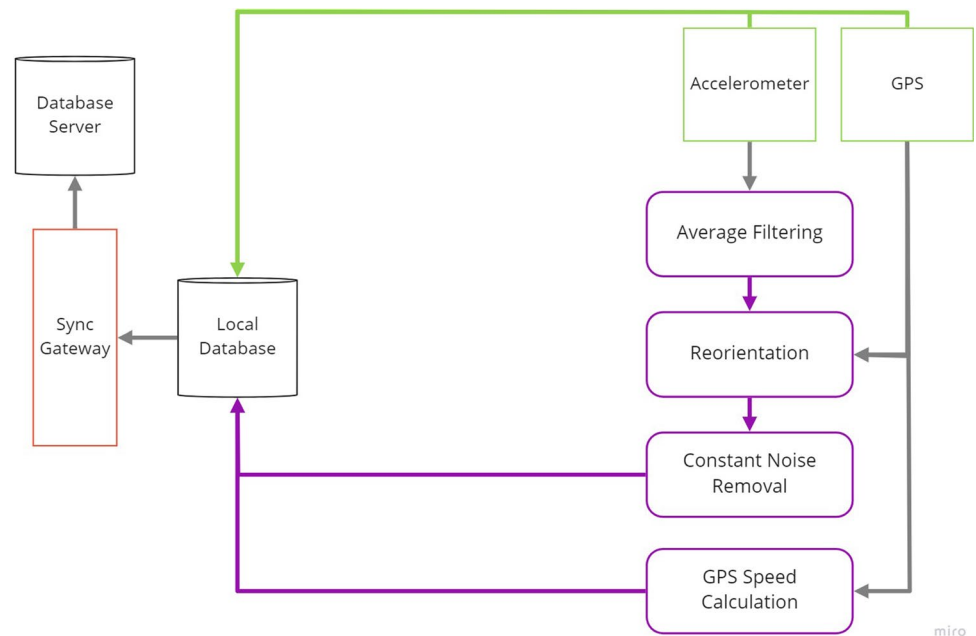


Fig. 3 The overall architecture of the iRoads X mobile application



the distance between each two GPS coordinates is calculated using Eq. 1.

The points mounting up to the desired segment size (500m) are saved as a new segment if the road did not change and a previously recorded segment did not overlap with the new segment before reaching the desired distance. The change of the road is detected using reverse geocoding. If the road changed, then a new segment will be started at that point discarding the points before that as shown in Fig. 4.

The overlapping with an existing segment can happen in two ways. The new journey can start at an earlier point in the road and then start to overlap with an old segment as shown in Fig. 5, or the new journey can start at the middle of an existing segment as shown in Fig. 6. In the former case, the existing segment is detected by searching for a starting coordinate of an already saved segment that lies close to the GPS coordinate of the current journey. In the

latter case, an ending coordinate will be searched. The flow chart of the searching function is given in Fig. 7. If such a segment exists, the points up to that overlapping point in the current journey will be discarded.

Then the old segment is compared with the new journey to confirm whether they overlap from the starting point to the ending point. This is done by checking if the difference of the bearing angles of the two journeys does not exceed a predefined limit. This comparison is required as the new journey can have a U-turn at a middle point of the old segment, without reaching the segment end. On rare occasions, the two journey directions can show a significant difference due to other reasons, but the two segments may still completely overlap. To handle such cases another function is applied. The difference of that function is that instead of checking the bearing angle of every sub-segment, it will check the distance between the

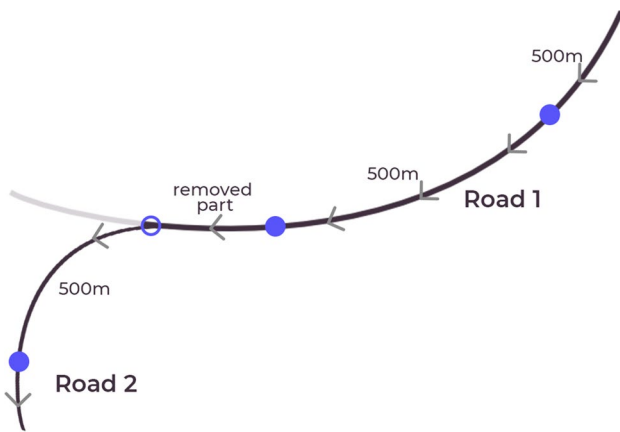


Fig. 4 Starting a new segment when the road changes

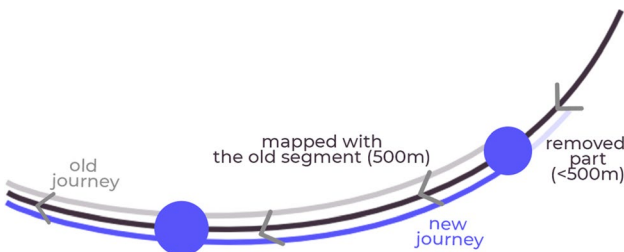


Fig. 5 New journey starting at an earlier point in the road and then overlapping with an old segment

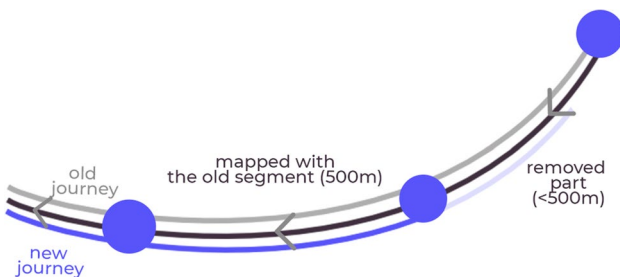


Fig. 6 New journey starting at the middle of an existing segment

ending point of the old segment and the new segment as shown in the flow chart in Fig. 8.

3.8 Road Roughness Prediction

When properly trained, machine learning models are capable of handling complex tasks and predicting more accurate values to a larger range of inputs compared to threshold-based algorithms [12, 26]. Therefore to predict road roughness level, an XGBoost based machine learning model is implemented. The predicted roughness level is

based on the IRI (International Roughness Index) and will be referred to as alt-IRI (alternative IRI) in the following sections. XGBoost which stands for “Extreme Gradient Boosting” is a highly scalable, memory efficient, and flexible supervised learning technique based on decision tree ensembles [11]. It attempts to predict an accurate value for the target variable by optimizing the loss function and applying regularization.

The features selected to predict the alt-IRI for each segment are the mean, the maximum, and the standard deviation of acceleration signals and the number of spikes (pulses) in x, y, and z directions, along with the average speed of the vehicle. The model is trained using a dataset collected from roads mounting up to 56 km through crowdsourcing. The dataset is labelled using the values collected by running a ROMDAS (Road Measurement Data Acquisition System) on the same road segments. The road quality category is then determined by classifying the predicted value into four classes as given in Table 2.

3.9 Updating the Values

If a new segment is identified, the predicted alt-IRI for the segment is saved along with the segment data and the timestamp. If the segment is repeated, then an updated value is calculated as given in 1.

Pseudocode 1 Calculating an updated mean.

```

1: function CALCULATE TRIMMED VALUE
2:    $NR \leftarrow$  number of records from the last 06 months
3:    $AR \leftarrow$  all records from the last 06 months
4:    $LR \leftarrow$  records where  $altIRI$  is among 5% of the largest  $altIRI$ s of all records in the last 06 months
5:    $SR \leftarrow$  records where  $altIRI$  is among 5% of the smallest  $altIRI$ s of all records in the last 06 months
6:   if  $NR \geq 25$  then
7:     remove  $LR$  and  $SR$  from  $AR$ 
8:      $SW \leftarrow 0$ 
9:      $SV \leftarrow 0$ 
10:    for remaining record in  $AR$  do
11:      calculate a weighted value for  $altIRI$  based on the recorded month
12:       $W \leftarrow$  assigned weight
13:       $V \leftarrow$  weighted  $altIRI$ 
14:       $SW \leftarrow SW + W$ 
15:       $SV \leftarrow SV + V$ 
16:    end for
17:     $trimmedmean \leftarrow SV/SW$ 
18:    return  $trimmedmean$ 
19:  end if
20: end function

```

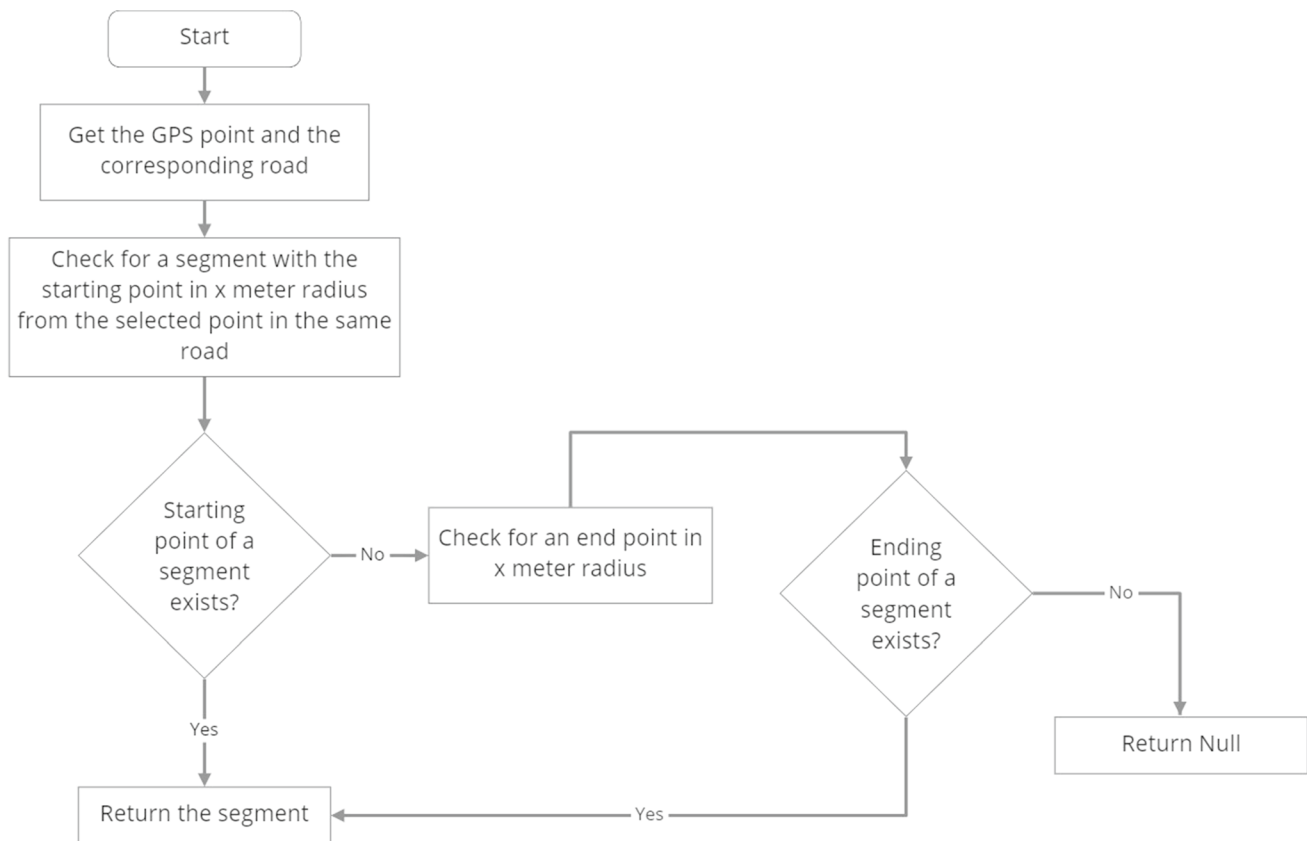


Fig. 7 The flow chart for the function to search existing matching segments

The updating algorithm ensures that the alt-IRI values are frequently updated while penalizing the possible errors due to crowdsourced data collection. If there exist parts of the roads with less than 500m of length that may not be included in a segment, then their alt-IRI values are calculated by averaging the values of adjacent segments.

3.10 Visualizing the Calculated Values

As the next step, a virtual dashboard is developed using AngularJS to visualize the collected alt-IRI of roads in Sri Lanka on a map. This dashboard can be used by the public and road development authorities to easily identify the road conditions in the country. Figure 9 presents the interface of the dashboard displaying calculated alt-IRI.

3.11 Overall System Overview

The complete system architecture is depicted in Fig. 10. This diagram depicts how the aforementioned components communicate with one another in order to achieve the final objective.

Figure 11 shows a conceptual diagram that explains the overall system. The 3 axis accelerometer readings are the independent variables in this system. The IRI of roads is the dependent variable. The moderator variables are the type of vehicle, the type of smartphone, and the positioning of the smartphone. These values are managed, by utilizing crowdsourcing techniques in such a way that the effect on the system's outcomes from these variables is reduced.

Table 3 compares the techniques utilized in this system to those used in existing similar systems. Based on the advantages and limitations of prior studies mentioned in Table 3, it can be concluded that the proposed techniques in this paper are more suitable for crowdsourcing compared to the previous approaches.

4 Experiments

To evaluate the developed system, a set of field tests were carried out. Field tests were arranged to analyze the effects of the variations in passenger car type, road condition,

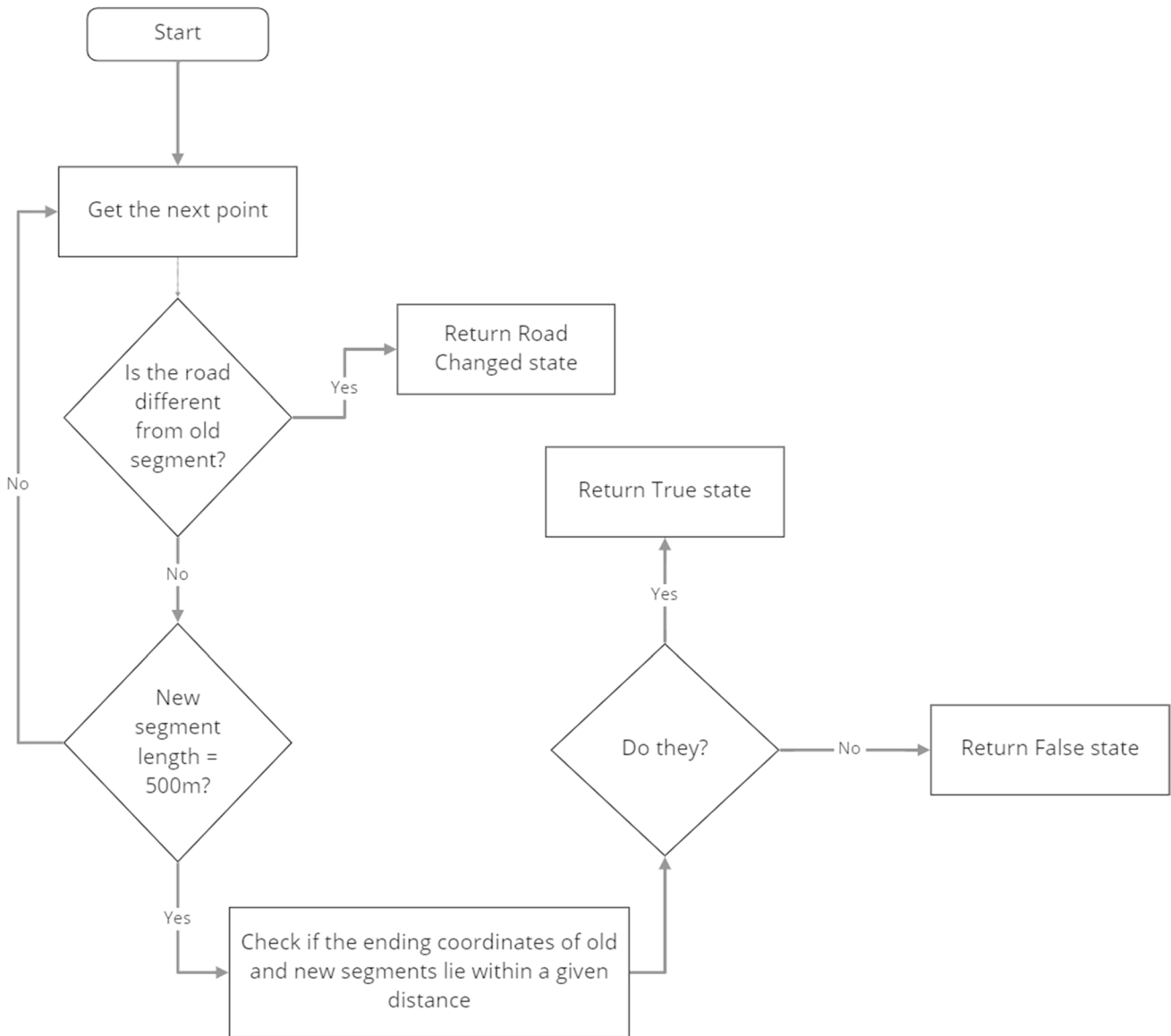


Fig. 8 The flowchart of the second algorithm to confirm that the new segment overlaps with the old segment

Table 2 The road quality category according to the alt-IRI

alt-IRI	Category
<3.5	Very Good
3.5 – 4.5	Good
4.5 – 6	Fair
>6	Poor

smartphone model, and the position of the smartphone inside the vehicle.

The car type, smartphone position, and smartphone model are chosen since the degree of noises, such as engine

vibrations, can vary depending on these factors. The smartphone orientation can also change based on the position of the smartphone. To analyze the impact of reorientation, noise filtering and roughness prediction techniques used in the proposed solution, experiments were carried under different controlled conditions.

In the first experiment, the data is collected on a selected road segment while the conditional factors vary. The selected segment is shown in Fig. 12. The vehicle type and smartphone placement are changed on each journey. Two journeys were conducted for each combination of vehicle type and smartphone placement. 3 vehicle types and 5 smartphone

iRoads X Data Analytics Platform

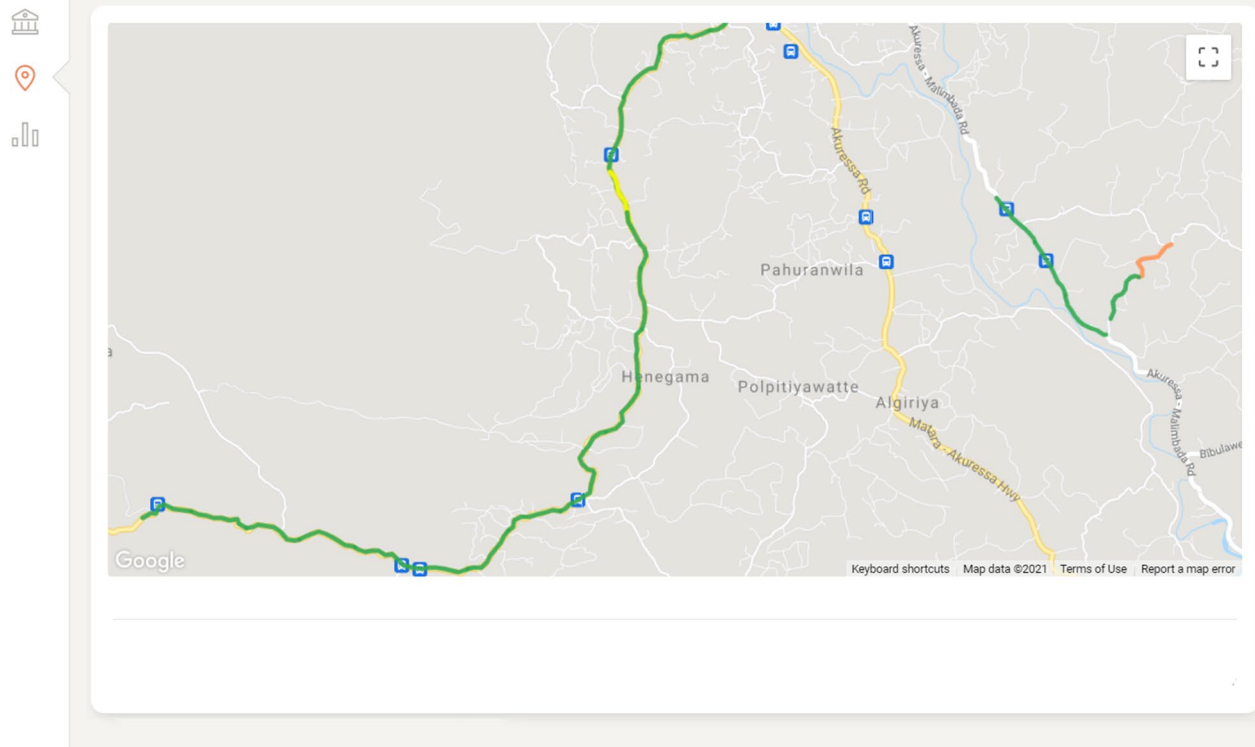


Fig. 9 The interface of the dashboard displaying calculated alt-IRI

Fig. 10 Overall architecture diagram of the system

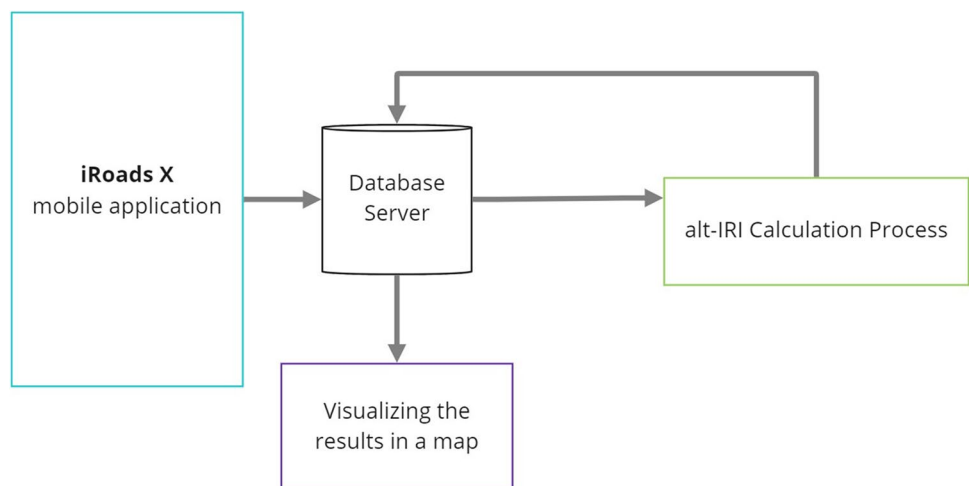
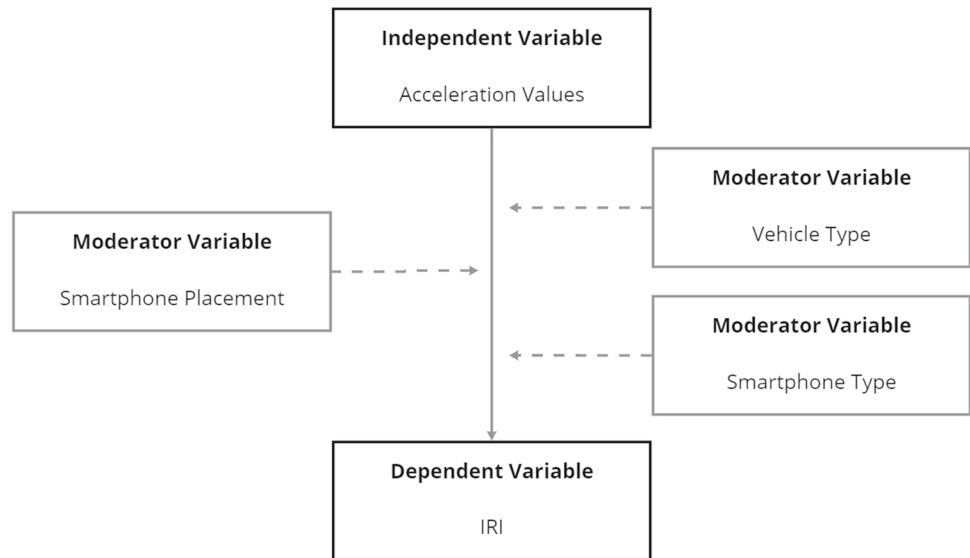


Fig. 11 Conceptual diagram of the overall system



placements were used and altogether 30 journeys were conducted for the selected road segment. The specifications of the selected vehicles are given in Table 4. In 10 journeys, the data were collected from 2 smartphone model types under identical conditions, i.e. in the same road segment, with the same smartphone placement and vehicle type. A summary of the experiment is given in Table 5.

The purpose of this experiment is to evaluate the system performance under different data collection environments. When collecting data by crowdsourcing it is necessary to ensure that the predicted values by the system are not affected by different conditional factors. The system is expected to provide similar roughness values under each different condition as the experiment is done on the same road segment.

Additionally, to evaluate the system performance under different road conditions, a set of roads under poor, fair, good and very good conditions were selected based on the visual inspection of experts. Then the alt-IRI values are predicted for each road segment using the developed system and subsequently, they are classified into the 4 classes based on the predicted value. The goal of this experiment is to evaluate the performance under different road conditions (very good, good, fair, poor) while the conditional factors are not controlled. Here the system is expected to predict different roughness values according to the road condition.

The results of these experiments are discussed in the following section.

5 Results and Discussion

The collected acceleration signals from each journey in the first experiment are analyzed to identify how the varying factors affect the collected data. Figure 13a and 13b show two example acceleration datasets collected under different conditions in the same road segment. Even though the raw acceleration signals demonstrate significant differences, after applying reorientation and noise filtering techniques the differences appear to be negligible.

Table 6 presents the predicted alt-IRI for the selected road segment. It appears that the predicted values do not have a significant variance across varying conditions. Therefore it can be concluded that passenger car type or smartphone placement differences have not imposed a significant effect on the final predictions albeit the differences shown in raw acceleration signals.

Table 7 summarizes the final predicted alt-IRI from the data collected from the two models Samsung SM-M022G (Model 1) and Samsung SM-A920F (Model 2) under identical conditions. Values predicted by the smartphone Model 2 seem to have slightly higher values compared to the values from Model 1 due to the differences in the sensitivity of the sensors, but when compared with all the predictions a significant variance cannot be observed.

The results suggest that the proposed techniques have successfully reduced the conditional effects. Removing the constant noise has helped in minimizing the effect of the car type and the smartphone type and applying

Table 3 A comparison of techniques used in this paper and in previous work

Features	Techniques	Advantages	Limitations
Noise filtering	[7, 22, 39] This paper	High pass filtering / Band pass filtering Simple moving average filtering and constant noise filtering.	May remove some vibrations that would carry useful information about pavement roughness
Reorientation	[9], This paper	Using smartphone, geometric and vehicle coordinate system	Requires a magnetometer but some smartphone models do not have magnetometers.
Automatic Journey Detection	[31, 39], This paper [25]	Using breaking events and Euler angles Using GPS Change	Can be used with many smartphone models False detection when walking or running
Automatic Route Segmentation	This paper [7, 33, 36] [20]	Using Pathsense SDK Based on time period Using predefined segments	May require slightly additional time to identify the journey after the journey has actually started Cannot merge multiple sources Need to segmentize all the roads manually
Database	[25] This paper [36]	Using rectangular polygons Proposed method in this paper PostgreSQL	Multiple roads can be included in the same segment. Require more computation time Need to implement other methods to collect data when the internet connection is not available Slow performance with large amounts of data
Roughness Value Prediction	This paper [19, 24, 25, 38]	Couchbase Lite Threshold and/or equation based	Does not support geospatial queries Results are easily affected by factors other than the actual road roughness Threshold values have to be adjusted, when adapting to different conditions

Table 3 (continued)

Features	Techniques	Advantages	Limitations
	[35]	Dynamic Time Warping (DTW) Capable of handling different vehicle speeds	Not applicable for calculating road roughness, better suited for road anomaly classification
	[22]	Fuzzy logic No training required	Rules in the fuzzy logic system need to be updated when adapting to different conditions
Merging Values from Multiple Sources [23]	[7, 21, 36], This paper	Machine learning models Predictions can be adjusted for different conditional factors	Requires model training
	[23]	Clustering approaches Provides more positional accuracy	Resource and time expensive to compute for road segments, better suited for road anomaly detection
	[21, 25, 38]	Calculating simple average Require less time for the calculation	Errors and outliers in predictions can affect the final value GPS errors can lead to positional inaccuracies
	This paper	Calculating trimmed mean weighted by timestamp Prediction outliers are removed providing a more accurate final value Comparable time complexity to calculating a simple average Possible changes in the roughness over time are captured	GPS errors can lead to positional inaccuracies GPS errors can lead to positional inaccuracies

Fig. 12 The selected road segment

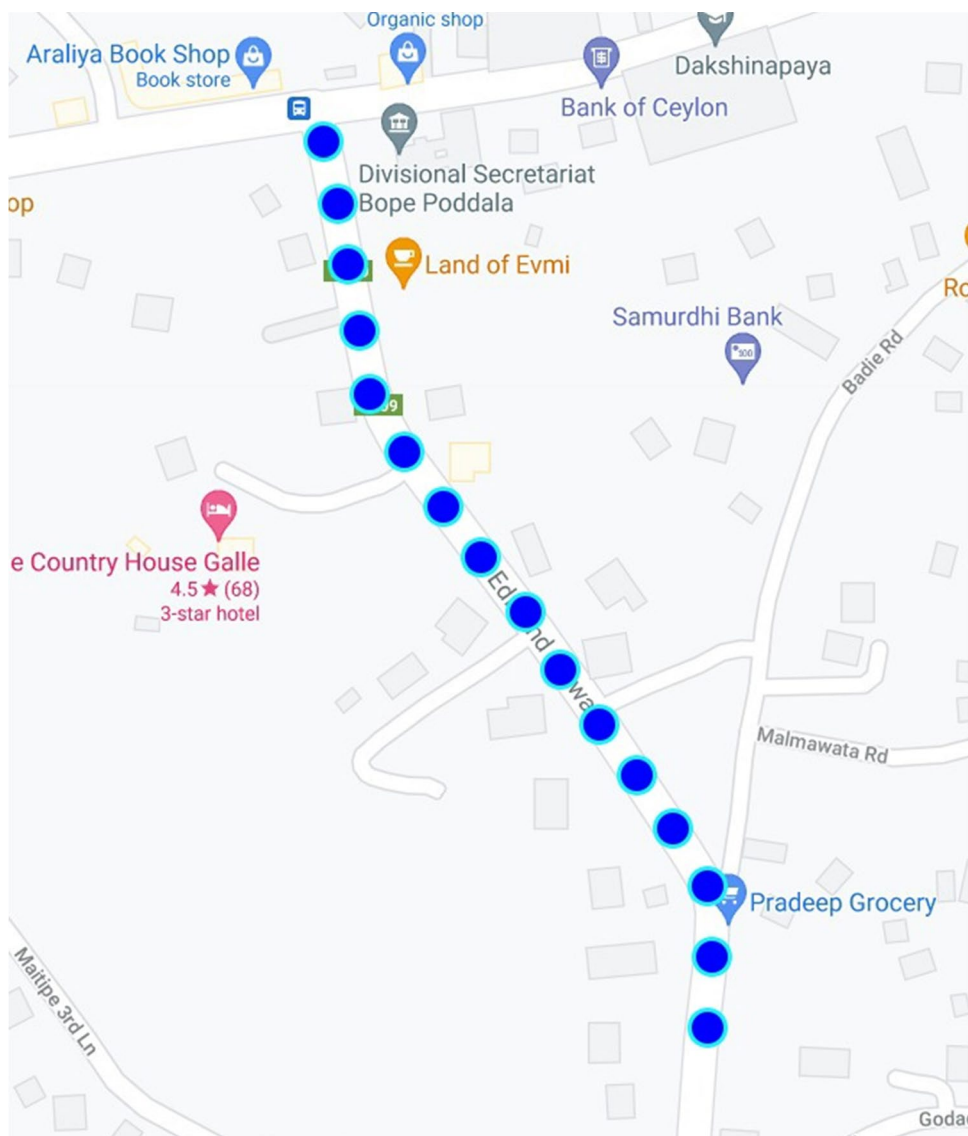


Table 4 The specifications of the selected vehicles

Passenger car brand	Model	Year
Toyota	Axio	2018
Mercedes Benz	C200	2019
Honda	CRV	2019

the simple moving average filtering has minimized the effects of hand movements and other high-frequency noises. The effect of the placement of the smartphone is reduced by applying the reorientation as it reorients the smartphone coordinate system to the vehicle coordinate

system regardless of the differences in the placement inside the vehicle.

After applying the updating algorithm to the predicted values for the segment in Fig. 12, an average value of 3.350 is achieved. Figure 14 demonstrates the value distribution, applied weights, and the removed values by the algorithm. It is evident from the results that the possible anomalous predictions are removed by the updating algorithm and a final averaged and more accurate value is obtained.

One of the reasons for the observed slight differences in the calculated values is that on different rides the vehicle does not drive on the same line and different lines on the same road may have slightly different roughnesses,

Table 5 A summary of the first experiment

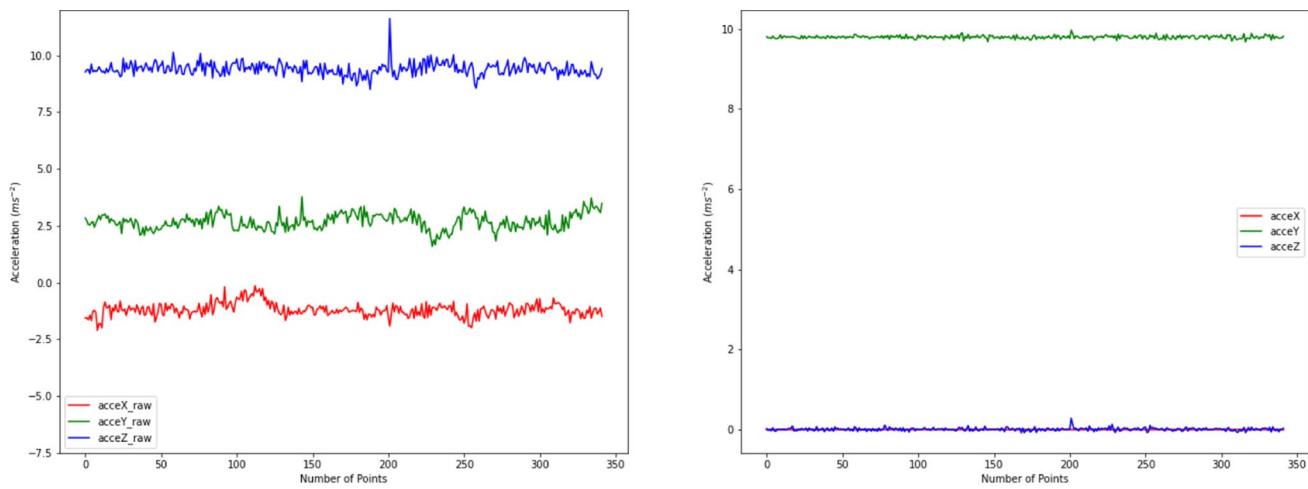
Journey	Vehicle	Smartphone Model	Position
1	Axio	Samsung SM-M022G	Near gearshift
2	C200	Samsung SM-M022G	Near gearshift
3	CRV	Samsung SM-M022G	Near gearshift
4	Axio	Samsung SM-M022G	On hand
5	C200	Samsung SM-M022G	On hand
6	CRV	Samsung SM-M022G	On hand
7	Axio	Samsung SM-M022G	On backseat
8	C200	Samsung SM-M022G	On backseat
9	CRV	Samsung SM-M022G, Samsung SM-A920F	On backseat
10	Axio	Samsung SM-M022G	Inside pants pockets
11	C200	Samsung SM-M022G	Inside pants pockets
12	CRV	Samsung SM-M022G, Samsung SM-A920F	Inside pants pockets
13	Axio	Samsung SM-M022G	Inside car door storage box
14	C200	Samsung SM-M022G	Inside car door storage box
15	CRV	Samsung SM-M022G, Samsung SM-A920F	Inside car door storage box
16	Axio	Samsung SM-M022G, Samsung SM-A920F	Near gearshift
17	C200	Samsung SM-M022G	Near gearshift
18	CRV	Samsung SM-M022G	Near gearshift
19	Axio	Samsung SM-M022G	On hand
20	C200	Samsung SM-M022G, Samsung SM-A920F	On hand
21	CRV	Samsung SM-M022G, Samsung SM-A920F	On hand
22	Axio	Samsung SM-M022G	On backseat
23	C200	Samsung SM-M022G, Samsung SM-A920F	On backseat
24	CRV	Samsung SM-M022G	On backseat
25	Axio	Samsung SM-M022G, Samsung SM-A920F	Inside pants pockets
26	C200	Samsung SM-M022G	Inside pants pockets
27	CRV	Samsung SM-M022G	Inside pants pockets
28	Axio	Samsung SM-M022G, Samsung SM-A920F	Inside car door storage box
29	C200	Samsung SM-M022G, Samsung SM-A920F	Inside car door storage box
30	CRV	Samsung SM-M022G	Inside car door storage box

but after averaging, the updated value is expected to contain the most desirable value for the complete road segment. Another factor that affects the z-axis acceleration is the vehicle speed [25]. Therefore the predicted alt-IRI values are further analyzed with the average vehicle speed to figure out if there is an influence of speed on the predicted alt-IRI. The obtained results are shown in Fig. 15. It can be observed that the two variables do not demonstrate a significant correlation. It is expected, as the average speed is one of the input features for the machine learning model.

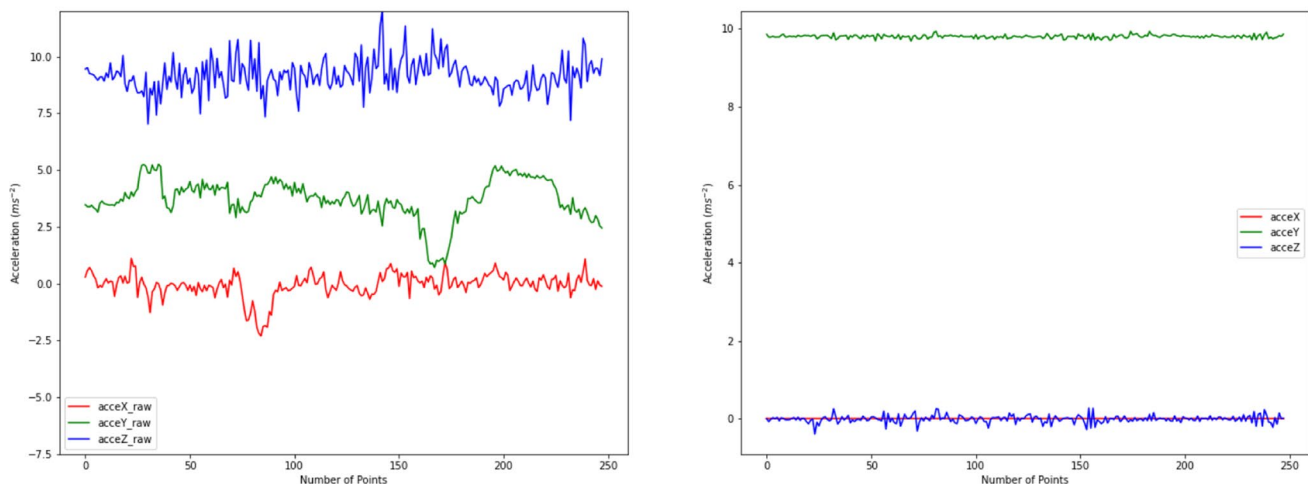
It can be observed that after applying the processing techniques, and the machine learning model predicting similar values for the segment when the segment changes and the roughness level changes the system must be able

to predict different alt-IRI values and the corresponding classes accordingly. The results obtained from the field tests of evaluating the system on road segments with different roughness levels proved the capability of the system as shown in Table 8. It can be observed that even after the acceleration signals have gone through the processing steps, the changes were captured and the machine learning model predicted the alt-IRI values accordingly.

The results from this second experiment highlight the advantage of using machine learning models to predict the IRI values as it gives the ability to predict accurate values for a larger range of complex inputs. Furthermore, the overall results prove that the segmentation algorithm accurately segmentizes the routes, identifies the overlapping



(a)



(b)

Fig. 13 Example acceleration signals collected from the first experiment (a) Raw and processed acceleration signals collected from Samsung SM-M022G placed near gearshift in C200, (b) Raw and

processed acceleration signals collected from Samsung SM-M022G placed on hand in C200

Table 6 The predicted alt-IRI for the selected road segment

Vehicle	Near gearshift		On hand		Backseat		Pocket		Car door	
	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd
C200	3.391	3.367	3.382	3.192	3.328	3.348	3.459	3.656	3.258	3.229
Axio	3.308	3.143	3.319	3.378	3.423	3.504	3.328	3.284	3.319	3.319
CRV	3.319	3.273	3.244	3.290	3.229	3.319	3.372	3.328	3.287	3.377

Table 7 The predicted alt-IRI from the data collected from the two smartphone models

Vehicle	Position	Predicted value	
		Model 1	Model 2
CRV	On hand	3.290	3.319
CRV	Pocket	3.372	3.551
CRV	Door	3.287	3.448
CRV	Backseat	3.229	3.323
Axio	Gearshift	3.143	3.198
Axio	Pocket	3.284	3.328
Axio	Door	3.319	3.379
C200	On hand	3.192	3.342
C200	Door	3.258	3.405
C200	Backseat	3.348	3.326

segment and enables the updating algorithm to calculate an averaged value.

The authors understand that the scope of the conducted field tests is limited, but the results achieved from the tests showed the proposed system’s capability to provide accurate predictions of the road conditions by crowdsourcing. However, as the target vehicle class of the developed system is passenger cars, the system is not trained or tested for other vehicle classes. The predictions may show some deviations

from the actual value if data is collected using a different vehicle class. Another major limitation in the proposed system is that there is a significant error rate of 10m in smartphone GPS, but many functions in the proposed solution such as the vehicle speed calculation and segmentation algorithm depend on the GPS coordinates. Furthermore, the accuracy of the segmentation algorithm depends on OpenStreetMap accuracy. If there are any unidentified roads or incorrectly named roads in OpenStreetMap, the route segmentation process may also be affected.

6 Conclusions and Future Work

This study attempts to develop a crowdsourced road surface quality monitoring application that requires minimum user interactions and puts minimum constraints on the user for collecting data, and which is also capable of reducing the effects of different factors that can deviate the calculated road roughness from the actual value. A set of experiments was conducted to verify the system performance under diverse conditions. The preliminary experiment was carried out on a selected route, using multiple car types, smartphone models, and different placements of the smartphone inside the vehicle. The obtained raw acceleration signals from the experiment showed considerable variations across conducted journeys, but after applying

Fig. 14 A visualization of the alt-IRI distribution, applied weights and the removed values by the updating algorithm

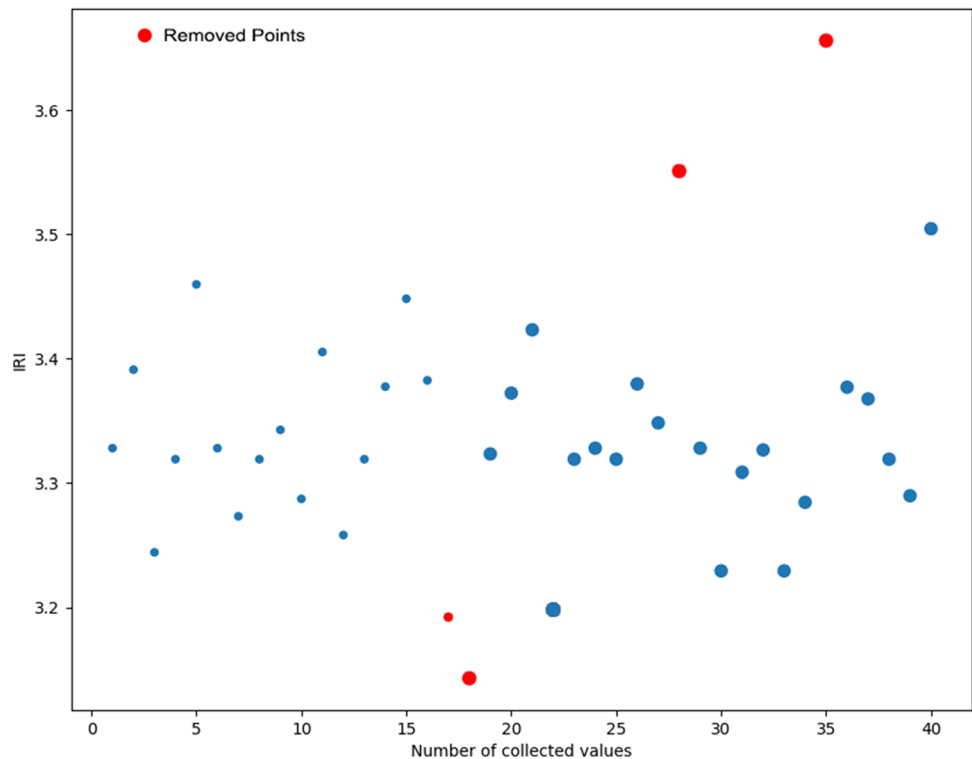
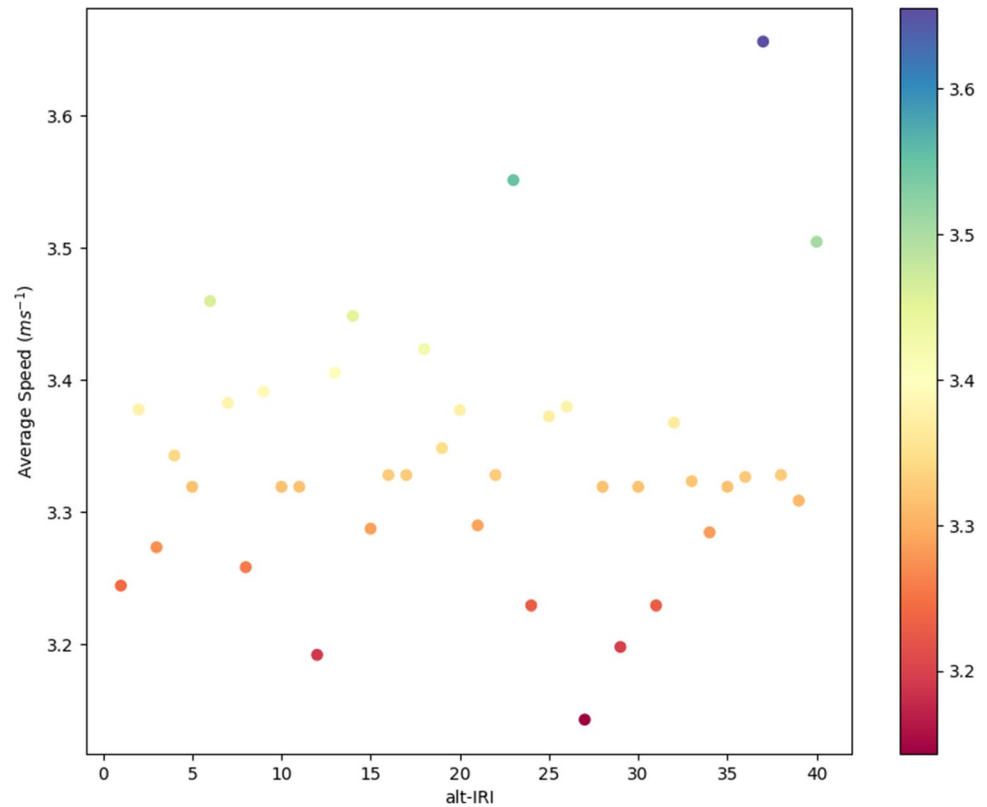


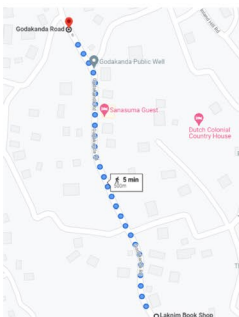
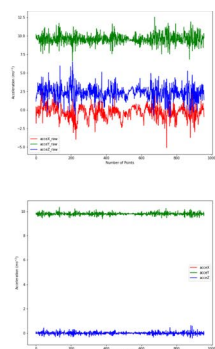
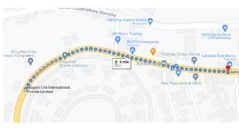
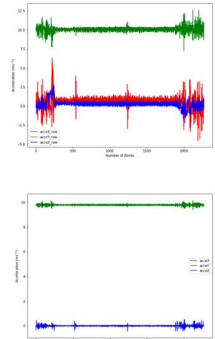
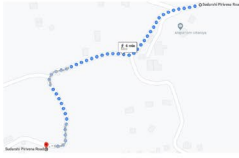
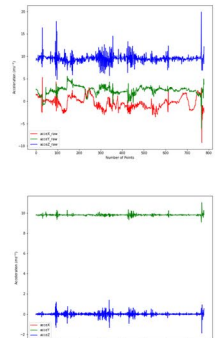
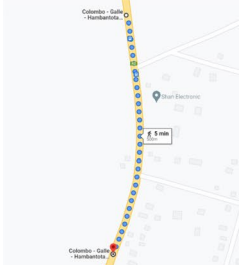
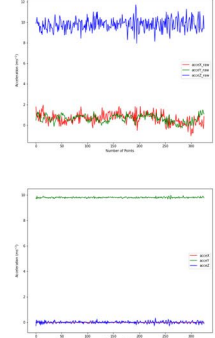
Fig. 15 The correlation between the average speed on the predicted alt-IRI



signal processing steps the deviations were minimized. Moreover, the developed machine learning model was able to predict accurate roughness levels further dealing with varying conditions. In the second experiment, the system was tested on different road conditions. It is observed that the system is capable of accurately determining the road quality category. When more data is collected frequently for the same road segment, it is observed that the final value obtained from the updating algorithm tends

to provide a more precise balanced value. Complexities can arise when thousands of mobile phone users connect to the system at the same time and a massive amount of data is being collected every second. Many recent studies have proposed efficient and scalable solutions for big data analysis and high-performance computing [10, 28]. As the proposed system is developed using the microservices architecture [13] it is possible to enhance the system performance by augmenting these techniques.

Table 8 The results obtained from the second field test

Road	Raw and processed acceleration signals	Predicted alt-IRI	Predicted class
<p>Godakanda road, Karapitiya (A 500m segment with a severely potholed surface)</p> 		6.372	Poor
<p>Avissawella Road, Homagama (A 500m segment with surface stripping)</p> 		4.743	Fair
<p>Sudarshi Pirivena Road, Akuressa (A 500m segment with few cracks on the surface)</p> 		3.975	Good
<p>Colombo-Galle-Hambantota-Wellawaya Road (A smooth 500m segment under very good condition)</p> 		3.143	Very good

Acknowledgements This research was supported by the Accelerating Higher Education Expansion and Development (AHEAD) Operation of the Ministry of Higher Education, Sri Lanka funded by the World Bank. Also, we acknowledge the support received from the Senate Research Committee Grant, University of Moratuwa, Sri Lanka.

Declarations

Conflict of Interests The authors declare that they have no conflict of interest.

References

- Couchbase lite. <https://www.couchbase.com/products/lite>
- Pathsense. <https://pathsense.com/>
- National Road Master Plan 2018–2028. Road development authority sri lanka (2018)
- Abeywardana, H., Abeywikrama, U., Amarasinghe, P.T., Kumarasinghe, R.: iroads - smartphone-based road condition monitoring
- Ahmed, H.U., Hu, L., Yang, X., Bridgelall, R., Huang, Y.: Effects of smartphone sensor variability in road roughness evaluation. *International Journal of Pavement Engineering*, pp. 1–6 (2021)
- Aleadelat, W., Ksaibati, K.: Estimation of pavement serviceability index through android-based smartphone application for local roads. *Transportation Research Record* **2639**(1), 129–135 (2017)
- Allouch, A., Koubaa, A., Abbes, T., Ammar, A.: Roadsense: Smartphone application to estimate road conditions using accelerometer and gyroscope. *IEEE Sensors J.* **17**(13), 4231–4238 (2017). <https://doi.org/10.1109/jsen.2017.2702739>
- Bapari, M., Haque, M., Chowdhury, D., Islam, M.J.: Impacts of unplanned urbanization on the socio-economic conditions and environment of pabna municipality bangladesh (2016)
- Bhoraskar, R., Vankadhara, N., Raman, B.: Kulkarni, P.: Wolverine: Traffic and road condition estimation using smartphone sensors. 2012 Fourth International Conference on Communication Systems and Networks (COMSNETS 2012). <https://doi.org/10.1109/comsnets.2012.6151382> (2012)
- Brdar, S., Gonzalez-Velez, H., Truiça, C.O., Benkner, S., Bajrovic, E., Papadopoulos, A., Novović, O., Grujić, N.: Big Data Processing, Analysis and Applications in Mobile Cellular Networks, pp. 163–185. https://doi.org/10.1007/978-3-030-16272-6_6 (2019)
- Chen, T., Guestrin, C.: Xgboost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*. <https://doi.org/10.1145/2939672.2939785>, pp 785–794. Association for Computing Machinery, New York, NY, USA (2016)
- Chugh, G., Bansal, D., Sofat, S.: Road condition detection using smartphone sensors: a survey. *International Journal of Electronic and Electrical Engineering* **7**(6), 595–602 (2014)
- Costa, D.I.C., Filho, E.P.e.S., Silva, R.F.d., de C. Quaresma Gama, T.D., Cortés, M.I.: Microservice architecture: A tertiary study. In: *Proceedings of the 14th Brazilian Symposium on Software Components, Architectures, and Reuse, SBCARS '20*. <https://doi.org/10.1145/3425269.3425277>, pp 61–70. Association for Computing Machinery, New York, NY, USA (2020)
- Dauni, P., Firdaus, M.D., Asfariani, R., Saputra, M.I.N., Hidayat, A.A., Zulfikar, W.B.: Implementation of haversine formula for school location tracking. *Journal of Physics: Conference Series* **1402**(7), 077028 (2019). <https://doi.org/10.1088/1742-6596/1402/7/077028>
- Dijkstra, L., Florczyk, A.J., Freire, S., Kemper, T., Melchiorri, M., Pesaresi, M., Schiavina, M.: Applying the degree of urbanisation to the globe: A new harmonised definition reveals a different picture of global urbanisation. *Journal of Urban Economics* **125**(103312), 103312 (2021)
- Douangphachanh, V., Oneyama, H.: A study on the use of smartphones under realistic settings to estimate road roughness condition EURASIP. *Journal on Wireless Communications and Networking* 2014(1). <https://doi.org/10.1186/1687-1499-2014-114> (2014)
- Du, Y., Liu, C., Wu, D., Jiang, S.: Measurement of international roughness index by using z-axis accelerometers and gps. *Math. Probl. Eng.* **2014**, 1–10 (2014). <https://doi.org/10.1155/2014/928980>
- Fontaras, G., Zacharof, N.G., Ciuffo, B.: Fuel consumption and co 2 emissions from passenger cars in europe – laboratory versus real-world emissions. *Prog. Energy Combust. Sci.* **60**, 97–131 (2017). <https://doi.org/10.1016/j.peccs.2016.12.004>
- Harikrishnan, P.M., Gopi, V.P.: Vehicle vibration signal processing for road surface monitoring. *IEEE Sensors J.* **17**(16), 5192–5197 (2017). <https://doi.org/10.1109/JSEN.2017.2719865>
- Islam, S., Buttlar, W.G., Aldunate, R.G., Vavrik, W.R.: Use of cellphone application to measure pavement roughness. T and DI Congress 2014. <https://doi.org/10.1061/9780784413586.053> (2014)
- Jones, H.: Roadroid continuous road condition monitoring with smart phones (2014)
- Kumar, R., Mukherjee, A., Singh, V.P.: Community sensor network for monitoring road roughness using smartphones. *Journal of Computing in Civil Engineering* **31**(3), 04016059 (2017). [https://doi.org/10.1061/\(asce\)cp.1943-5487.0000624](https://doi.org/10.1061/(asce)cp.1943-5487.0000624)
- Li, H., Goldberg, C., Yin, H.: Embracing crowdsensing: An enhanced mobile sensing solution for road anomaly detection. *ISPRS International Journal of Geo-information* **8**(9), 412 (2019). <https://doi.org/10.3390/ijgi8090412>
- Li, X., Goldberg, D.W.: Toward a mobile crowdsensing system for road surface assessment. *Comput. Environ. Urban. Syst.* **69**, 51–62 (2018). <https://doi.org/10.1016/j.compenvurbysys.2017.12.005>
- Lima, L.C., Amorim, V.J.P., Pereira, I.M., Ribeiro, F.N., Oliveira, R.A.R.: Using crowdsourcing techniques and mobile devices for asphaltic pavement quality recognition. In: *2016 VI Brazilian Symposium on Computing Systems Engineering (SBESC)*, pp. 144–149. <https://doi.org/10.1109/SBESC.2016.029> (2016)
- Lum, P.S., Shu, L., Bochniewicz, E.M., Tran, T., Chang, L.C., Barth, J., Dromerick, A.W.: Improving accelerometry-based measurement of functional use of the upper extremity after stroke: Machine learning versus counts threshold method. *Neurorehabilitation and Neural Repair* **34**(12), 1078–1087 (2020)
- Luu, H.N., Nguyen, N.M., Ho, H.H., Tien, D.N.: Infrastructure and economic development in developing economies. *Int. J. Soc. Econ.* **46**(4), 581–594 (2019). <https://doi.org/10.1108/ijse-05-2018-0252>
- Álvarez Cid-Fuentes, J., Álvarez, P., Amela, R., Ishii, K., Morizawa, R.K., Badia, R.M.: Efficient development of high performance data analytics in python. *Future Generation Computer Systems* **111**, 570–581 (2020). <https://doi.org/10.1016/j.future.2019.09.051>, <https://www.sciencedirect.com/science/article/pii/S0167739X18321393>
- Medina, J.R., Salim, R., Underwood, B.S., Kaloush, K.: Experimental study for crowdsourced ride quality index estimation using smartphones. *Journal of Transportation Engineering, Part B: Pavements* **146**(4), 04020070 (2020). <https://doi.org/10.1061/jpeodx.0000225>
- Meijer, A., Bolívar, M.P.R.: Governing the smart city: a review of the literature on smart urban governance. *Int. Rev. Adm. Sci.*

- 82(2), 392–408 (2016). <https://doi.org/10.1177/0020852314564308>
31. Mohan, P., Padmanabhan, V.N., Ramjee, R.: Nericell: Using mobile smartphones for rich monitoring of road and traffic conditions. In: Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems, SenSys '08. <https://doi.org/10.1145/1460412.1460450>, pp 357–358. Association for Computing Machinery, New York, NY, USA (2008)
32. Nirmal, P., Disanayaka, I., Haputhanthri, D., Wijayasiri, A.: Transportation mode detection using crowdsourced smartphone data. In: 2021 28Th Conference of Open Innovations Association (FRUCT), pp. 341–349. <https://doi.org/10.23919/FRUCT50888.2021.9347625> (2021)
33. Perttunen, M., Mazhelis, O., Cong, F., Kauppila, M., Leppänen, T., Kantola, J., Collin, J., Pirttikangas, S., Haverinen, J., Ristaniemi, T., et al.: Distributed road surface condition monitoring using mobile phones. *Ubiquitous Intelligence and Computing Lecture Notes in Computer Science*, pp. 64–78. https://doi.org/10.1007/978-3-642-23641-9_8 (2011)
34. Sattar, S., Li, S., Chapman, M.A.: Road surface monitoring using smartphone sensors: A review. <https://doi.org/10.32920/14638491.v1> (2021)
35. Singh, G., Bansal, D., Sofat, S., Aggarwal, N.: Smart patrolling: An efficient road surface monitoring using smartphone sensors and crowdsourcing. *Pervasive and Mobile Computing* **40**, 71–88 (2017). <https://doi.org/10.1016/j.pmcj.2017.06.002>, <https://www.semanticscholar.org/paper/0af1a90cc44b6a24c26e707e756397c5d161e4fa>
36. Souza, V.M., Giusti, R., Batista, A.J.: Asfalt: a low-cost system to evaluate pavement conditions in real-time using smartphones and machine learning. *Pervasive and Mobile Computing* **51**, 121–137 (2018). <https://doi.org/10.1016/j.pmcj.2018.10.008>
37. Staniek, M.: Repeatability of road pavement condition assessment based on three-dimensional analysis of linear accelerations of vehicles. *IOP Conference Series: Materials Science and Engineering* **356**, 012021 (2018). <https://doi.org/10.1088/1757-899x/356/1/012021>
38. Staniek, M.: Road pavement condition diagnostics using smartphone-based data crowdsourcing in smart cities. *Journal of Traffic and Transportation Engineering (English Edition)* **8**(4), 554–567 (2021). <https://doi.org/10.1016/j.jtte.2020.09.004>
39. Vittorio, A., Rosolino, V., Teresa, I., Vittoria, C.M., Vincenzo, P.G., Francesco, D.M.: Automated sensing system for monitoring of road surface quality by mobile devices. *Procedia, Social and Behavioral Sciences* **111**, 242–251 (2014). <https://doi.org/10.1016/j.sbspro.2014.01.057>, <https://www.sciencedirect.com/science/article/pii/S1877042814000585>
40. Wang, G., Burrow, M., Ghataora, G.: Study of the factors affecting road roughness measurement using smartphones. *Journal of Infrastructure Systems* **26**(3), 04020020 (2020)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.