A participatory sensing framework to classify road surface quality

Davidson E. Nunes[1,3] and Vinicius F. S. Mota[2]*

Abstract

Participatory sensing networks rely on gathering personal data from mobile devices to infer global knowledge. Participatory sensing has been used for real-time traffic monitoring, where the global traffic conditions are based on information provided by individual devices. However, fewer initiatives address asphalt quality conditions, which is an essential aspect of the route decision process. This article proposes Streetcheck, a framework to classify road surface quality through participatory sensing. Streetcheck gathers mobile devices' sensors such as Global Positioning System (GPS) and accelerometer, as well as users' ratings on road surface quality. A classification system aggregates the data, filters them, and extracts a set of features as input for supervised learning algorithms. Twenty volunteers carried out tests using Streetcheck on 1,200 km of urban roads of Minas Gerais (Brazil). Streetcheck reached up to 90.64% of accuracy on classifying road surface quality.

Keywords: Participatory sensing, Road surface quality, Supervised learning, Crowdsensing

1 Introduction

Streets and roads are worldwide daily popular routes for transportation, and yet, drivers face road surface conditions problems in several countries. For instance, low-quality street asphalt is one of the leading causes of car accidents in Brazil [1].

Today, smart devices allow drivers to get real-time traffic information and adapt their routes. To achieve accurate information, users collect and share data individually, and in exchange, they get useful aggregated information. This collective collaboration enables a participatory sensing network [2], also called crowdsensing [3]. Indeed, due to the popularity of smart devices, people are becoming "mobile sensors" capable of gathering valuable data about their environment [4].

Services such as *Waze*[1] and *MapLink*[2] aggregate data from personal devices and provide real-time traffic information for all users. Besides traffic conditions, road surface quality, potholes, manhole covers, and road humps on streets increase vehicles' fuel consumption and reduce passengers' comfort. However, the classification of road surface quality based on such irregularities poses open

challenges, such as a common definition of quality, the accuracy of sensors, and privacy. Despite several initiatives for real-time traffic monitoring, few of them consider road surface quality information, which may be crucial for route decision making.

Typically, road surface condition is given by the international roughness index (IRI) [5]. The index is computed from the longitudinal profile measurements using a quarter car simulation at a speed of 80 km/h [6]. Mapping the road conditions into an IRI requires specific equipments, such as laser profilers or high-frequency sample accelerometers, which are costly and demand skilled labor. These types of equipment must be calibrated and mounted on specific vehicles. Therefore, regular drivers are unable to calculate the IRI by using their smartphones. In fact, drivers are just required to know the overall quality of the pavement instead of the IRI.

To fulfill this context, this article presents *Streetcheck*, a participatory sensing framework to classify urban road surface based on data gathered and shared by users through their devices. Afterward, *Streetcheck* aggregates, filters, and classifies users' data to provide a map view of the classified routes.

Streetcheck is composed of a mobile application and an offline classification system. The mobile App gath-

---

*Correspondence: vinícius.mota@inf.ufes.br
[2]Department of Informatics, Universidade Federal do Espírito Santo, Vitória, Brazil
Full list of author information is available at the end of the article

ers GPS sensor, which provides geographic positions and speed, and the three axes (x, y, z) gravitational acceleration measured by the device, commonly called as accelerometer sensor. Whenever possible, passengers rate their feeling about the road surface quality and provide other information, such as the type of vehicle and position of the device before submitting the data. The mobile application forwards the information to the classification system, which aggregates, filters and classifies the raw data by using supervised learning algorithms.

The accelerometer, built-in the device, can capture all vibrations caused by the roughness of the road surface to the vehicle. However, the primary challenge in measuring these vibrations should consider the position of the device inside the car, which could be on the passenger's hand, inside the glove compartment, at the dashboard, among others. Moreover, the suspension system of the car reflects on the accelerometers' readings.

To overcome the issues above, accelerometers' readings were mapped into a set of features to feed supervised learning algorithms, such as Random Forest, SVM, KNN, and J48. To evaluate the performance of the proposed features, twenty volunteers used Streetcheck while traveling by car (as drivers or as passengers) on 1.218,34 km for three weeks. The contributions of this work are summarized as follows:

- It proposes a Framework to gather devices sensors through participatory sensing for further filtering, classification, and visualization of the road conditions;
- It proposes a set of features used as input for supervised learning algorithms. These features overcome the device location issue mentioned above.
- It provides a proof of concept with relevant amount of data. During our experiments, *Streetcheck* gathered GPS and accelerometer data of users' devices in more than 1,000 km. Streetcheck extracts the features using these data.

The benefits must be higher than costs to engage users to cooperate in sensing and data transmission [7]. In this work, we considered the benefits of the aggregated information worth the collaboration, since route decision-making must consider road surface quality, in addition to the traffic condition.

The rest of this article is organized as follows: Section 2 provides a background and discusses the related work. Section 3 introduces *Streetcheck*, with the framework architecture and the features extracted from the raw data. Section 4 discusses the methodology of the tests and summarizes the data gathered. Section 5 presents the results and Section 6 concludes and discusses the future research on the matter.

2 Background

2.1 Machine learning techniques

Machine learning (ML) techniques have been used to infer knowledge from a given data set [8]. ML attempts to find patterns inside a labeled dataset, called the training dataset, to create a model able to identify similar patterns in unlabeled datasets. These ML techniques have been employed in several domains, from image recognition to networking traffic pattern [9].

Algorithms that learn from data can be classified into four learning paradigms: *supervised*, which uses labeled data as the training dataset to create the classification model; *semi-supervised*, when the training dataset is partially labeled; *unsupervised*, the one that creates the model based on clusters of unlabeled dataset; and *reinforcement learning*, which learns from the model based on examples from training datasets. Interested readers must refer to [8].

This work considered the supervised learning paradigm due to the following reasons: i) We allowed non-expert people rate their perception on the asphalt quality; ii) Therefore, different people may have different perceptions about the road surface quality; and iii) there are no clear clusters among accelerometers' raw data, which means there are no clear examples of good or bad pavements.

We have evaluated the performance of the following supervised learning algorithms to classify the Streetcheck dataset:

Random Forest: It builds a multitude of decision trees series during training and it classifies data for each of the generated trees. The classification is based on the rule of the majority, in which, each sample is assigned to the most frequent label among the classifiers.

SVM: It generates a multidimensional hyperplane that best separates a data set into two groups. In order to do so, it creates a model from a training set that contains labeled data.

KNN: It classifies the samples based on the neighbors' samples class. Each sample calculates its distance to the nearest K training samples. Then, the sample assigns itself as the same predominant class among its k-neighbors.

J48: It creates a classification tree based on a training dataset. Each node analyzes the relevance of a feature and selects those features that best divide (clustering) the dataset.

2.2 Road surface conditions based on accelerometers readings

Road surface roughness is based on visual inspections or using specialized instruments to take physical measurements of the road irregularities [10–12]. The estimate of the roughness index (IRI) requires a fixed accelerometer

in the car cabin, calibration to take into account the tires and the suspension system of the car, and skilled labor.

Accelerometer sensors and Global Positioning System (GPS) have been widely employed for detecting surface conditions [10, 11, 13–17]. In common, these proposals attempt to detect single anomalies such as potholes, bumps, or other road surface anomalies. The primary challenge faced by those previous researches concerns on mapping accelerometer raw data into an anomaly in the road.

Gonzales et al. used high-frequency sample accelerometer fixed in a specific vehicle type to estimate road conditions [10]. *Pothole Patrol System* (P2) is one of the first attempts to detect anomalies in the road surface [13]. Authors built devices equipped with a high-frequency sample accelerometer and GPS. Jakob et al. mounted these devices on the roof of seven taxis in Los Angeles. P2 applies Naive algorithm to detect the irregularities. A similar approach appears in [11]. Mednis et al. considered peaks in the raw accelerometer data to detect anomalies [11]. Their proposal achieves a hit rate up to 92% in anomalies detection. Furthermore, the accelerometers used in these previous works had higher sample frequency than built-in smartphones' accelerometers.

In [18], authors equipped vehicles with cameras and GPS for further classification of the images. This approach requires to label the images containing anomalies in the road surface. Afterward, the labeled images were used as training dataset for a SVM classifier. In [12], authors attached Arduino with an accelerometer in a fixed position of the vehicle. Furthermore, authors attempt to collect data at a constant speed and through roads with clear different conditions. Based on this data, they used machine learning algorithms with statistical models as features [12].

The accelerometer data is compared to the IRI measured on the same road in [19]. Authors filtered the samples to reduce the effect of accelerations caused by gravity and vehicles (breaks and acceleration) and the centrifugal accelerations at curves. In [15], authors estimate the IRI using smartphone accelerometer samples, segmented on 1s-length frames. Authors proposed a theoretical model, which maps the vertical acceleration of the samples into an IRI based on the speed and suspension model of the vehicles.

The mean-absolute-value of the vertical acceleration for every 100 meters was used to estimate the road conditions in [6]. Authors validated their proposal by attaching a smartphone on a motorbike in the northern region of India. In these previous proposals, accelerometers must be fixed in the vehicle to consider only one of the accelerometer axis, usually the gravity acceleration (Z-axis).

In [20], authors attached smartphones in the car panel to capture images, location, and accelerometer data. They extracted a set of features based on data gathered and applied random forest classifier to detect road conditions, such as the existence of dry snow or solid ice, bumps, and potholes. The "Bag of Words" approach, as proposed in [17], creates a feature vector as input to supervised learning algorithms. Measurements were made with a smartphone at several positions inside the vehicle to verify the influence of a given position upon the results.

Smart patrolling is another proposal that uses smartphones to gather accelerometer and GPS data from several users [21]. However, smart patrolling focused on increasing the accuracy for detecting potholes and bumps using Dynamic Time Warping (DTW) technique. Authors showed that different smartphones provide different results due to the quality and calibration of their sensors.

The research articles mentioned above focused on the detection of abnormalities, like holes, humps or path junctions. Brazilian streets, however, present particularities as ripples, which may go unnoticed and, therefore, influence the performance of these tools.

A crowdsourcing based road surface monitoring system, called CRSM, was proposed by [22]. CRSM is a hardware module containing accelerometer and GPS devices. CRSM applies a modified version of the Gaussian Mixture Model (GMM) to detect bumps and potholes. To evaluate CRSM, authors attached their hardware in the dashboard of 100 taxis in China.

A participatory sensing framework to evaluate road conditions is proposed in [23]. The authors propose an anomaly index based on the vectorial sum of the accelerometer axes. Therefore, the anomaly index ranks the road conditions. Streetcheck differs from [23] due to the ratings about the road conditions provided by users. A similar solution, RoadMonitor, as proposed by [24], used the gyroscope of smartphones to detect anomalies in the road. Authors cross-validated gyroscope and accelerometer to improve the accuracy of RoadMonitor.

Roadscan, as proposed by [16], classifies segments of streets based on the standard deviation and peaks of the accelerometer readings. Authors show the speed of the vehicle impacts on the accelerometer readings and thus the road roughness sensing.

In summary, the first solutions tackling road surface classification focused on detect anomalies, such as bumps and potholes, and they required specialized equipment [10, 11, 13]. Solutions aiming to map the road conditions into a metric appeared in [12, 15, 16, 20, 22, 25]. Among these solutions, only [16, 20] considered participatory sensing with users' collaboration. Table 1 summarizes the main characteristics of each one of the proposed solution.

Table 1 Summary of related work

Ref.	Only asphalt anomalies	Road surface conditions	Participatory sensing	Requires specialized equipments or skill	Year
[10]	YES	NO	NO	YES	2008
[13]	YES	NO	NO	YES	2008
[11]	YES	NO	NO	YES	2011
[22]	NO	YES	YES	YES	2013
[15]	NO	YES	NO	NO	2015
[23]	YES	NO	YES	NO	2015
[24]	YES	NO	NO	NO	2015
[18]	YES	NO	NO	NO	2016
[16]	NO	YES	YES	NO	2016
[21]	YES	NO	YES	NO	2017
[17]	YES	NO	NO	NO	2017
[20]	NO	YES	YES	NO	2017
[25]	NO	YES	NO	YES	2018
[12]	NO	YES	NO	YES	2019
Street Check	NO	YES	YES	NO	2019

This article proposes a participatory sensing framework, called StreetCheck, to classify road surface conditions based on users' ratings. To achieve this goal, Streetcheck is a framework composed of a mobile application, a cloud-based service, and an offline data analyzer. The mobile app allows users to sense and to rate the conditions of road surface stretches, in a participatory fashion. StreetCheck considers that casual drivers may have opinions about the road conditions, even when not being aware of the IRI metric. Furthermore, different drivers may have conflicting opinions about the condition of the same stretch of road.

Streetcheck maps the raw accelerometer data into a set of features and investigated the accuracy of several supervised learning algorithms to classify the unlabeled data. As a result, users have a global view of the road surface quality covered by all Streetcheck users.

3 Streetcheck architecture
3.1 General overview
The *Streetcheck* framework is composed of two modules: 1 **Data sourcing**, an application that runs on mobile devices; and 2 **A Classification system**, implemented on the cloud, responsible for receiving and storing data, and subsequently, data filtering, features extraction, and classification of the data. The overall architecture of *Streetcheck* is shown in Fig. 1.

Streetcheck gathers geolocation and accelerometer raw data, and optionally, a user can rate the roughness of the road. This rating is based on the user's perception of the road conditions. Each user's rating corresponds to

a 50-meter stretch. We define each 50m of the stretch of the road as a *sample*[3]. For each sample, local processing summarizes data, removes wrong readings, and bundles the raw data into a packet to transmit it to the classification system.

After receiving the raw data, the classification system filters them and extracts a set of features, as discussed in Section 3.3. The samples labeled by users are defined as the training dataset and feed the supervised learning algorithms. The non-labeled samples are submitted to the prediction process. A map provides a global view of the quality of the stretches of streets that *Streetcheck* users have already passed on.

3.2 Data sourcing on mobile devices
Streetcheck data sourcing runs on users' mobile devices equipped with a GPS and an accelerometer. As proof of concept, we have implemented Streetcheck for Android OS devices[4]. Streetcheck stores the accelerometer's reading, geolocation, and user's rating data on the device's memory and afterwards transfers the data to the classification system for further processing. Figure 2 presents Streetcheck's Graphical User Interface (GUI).

A Built-in Geolocation device provides the latitude, longitude, altitude positions, as well as the current speed of the device. *Streetcheck* stores the vehicle geographical location every 50 meters. Each user rating becomes a label for a sample. Figure 3 illustrates the sample labeling. Due to safety issues, *Streetcheck* alerts users to provide information only when they are passengers during a ride.
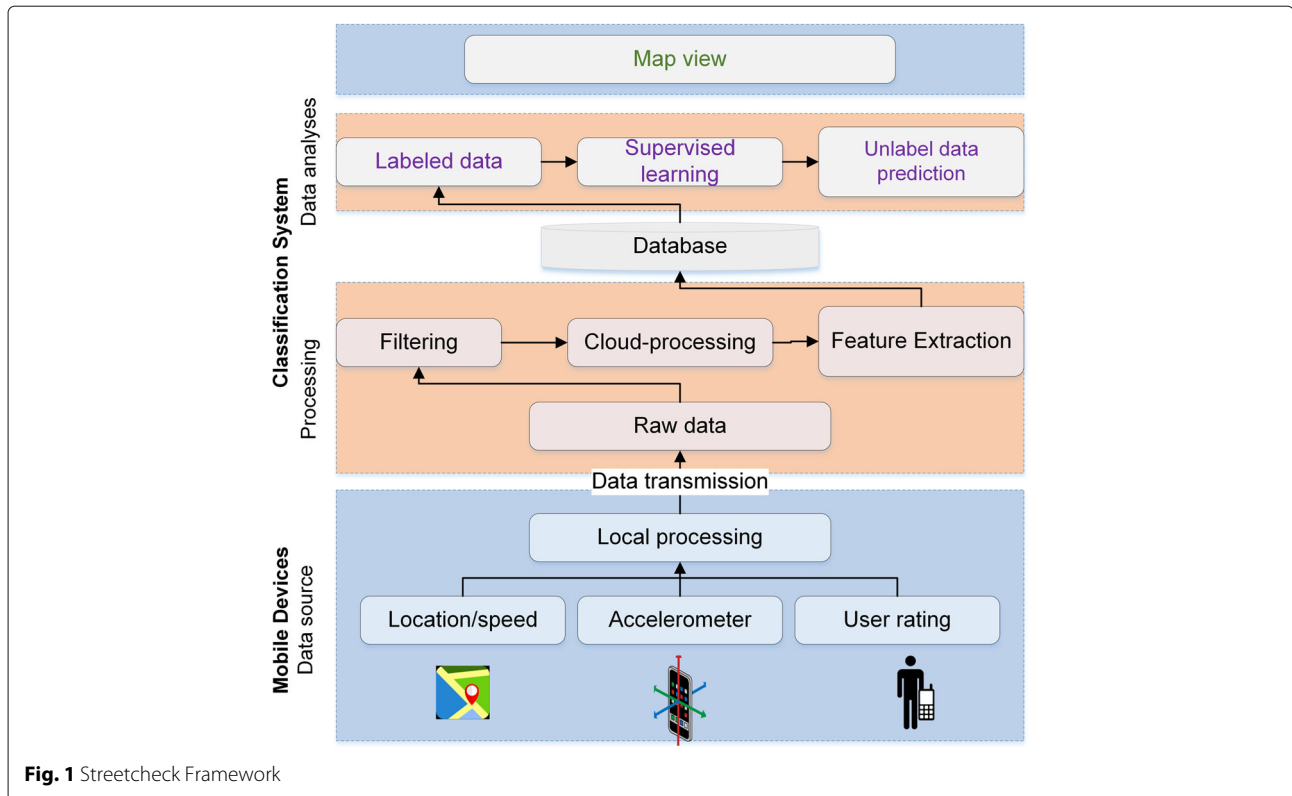
Fig. 1 Streetcheck Framework

Streetcheck allows users to rate stretches of streets in a binary mode, namely "GOOD" or "BAD". The binary labeling was defined after preliminary tests indicate that the majority of samples in the training dataset formed only two clusters. When a user rates a GOOD or BAD rating, the last sample (50-m of road stretch) is labeled with the correspondent rating. The classification system analyzes and classifies these samples.

One may note that the higher the speed, the higher the frequency that users would have to rate each 50-m of road stretch. Users must rate only when they observe a good or bad asphalt quality. Although increasing user engagement to assess the road conditions is out of the scope of this work, assuming there are several users in the same region, StreetCheck would require just a few entries of each user to have enough samples to be used in the training dataset.

Accelerometers measure the inertial forces along one, two or three axes. Smart devices are usually equipped with a three-axes gravitational acceleration sensor, which can be represented with a 3-dimensional Cartesian frame (x, y, z) [26]. The accelerometer captures the vibrations generated by the roughness of the road surface inside a moving vehicle.

The Accelerometers used to measure IRI have a higher frequency than the device's accelerometers. Indeed, device's accelerometers can perform 160 readings per second on average against 360Hz of specialized accelerometers [15]. In both cases, accelerometers generate a large amount of continuous data. Due to this reason, local processing summarizes these readings as a set of features for every 50 meters traveled (each sample of the street). Table 2 summarizes information stored for each sample. Each feature is described in the next section.

Since Streetcheck is a collaborative tool and can be used by non-experts in road surface conditions, different users can provide different ratings about the same stretch of road. To tackle this issue, the classification system clusters and filters the labeled dataset for supervised learning algorithms.

3.3 Classification system

After receiving the raw data, the classification system first applies a filter to remove incomplete samples and summarize the dataset. Next, the cloud-processing module extracts a set of features from the accelerometer readings and splits the dataset into labeled and unlabeled samples. The labeled samples become the training dataset for the following supervised learning algorithms: Random Forest, SVM, KNN, and J48. Based on the model created by these algorithms, the classification system infers the quality of the unlabeled samples.
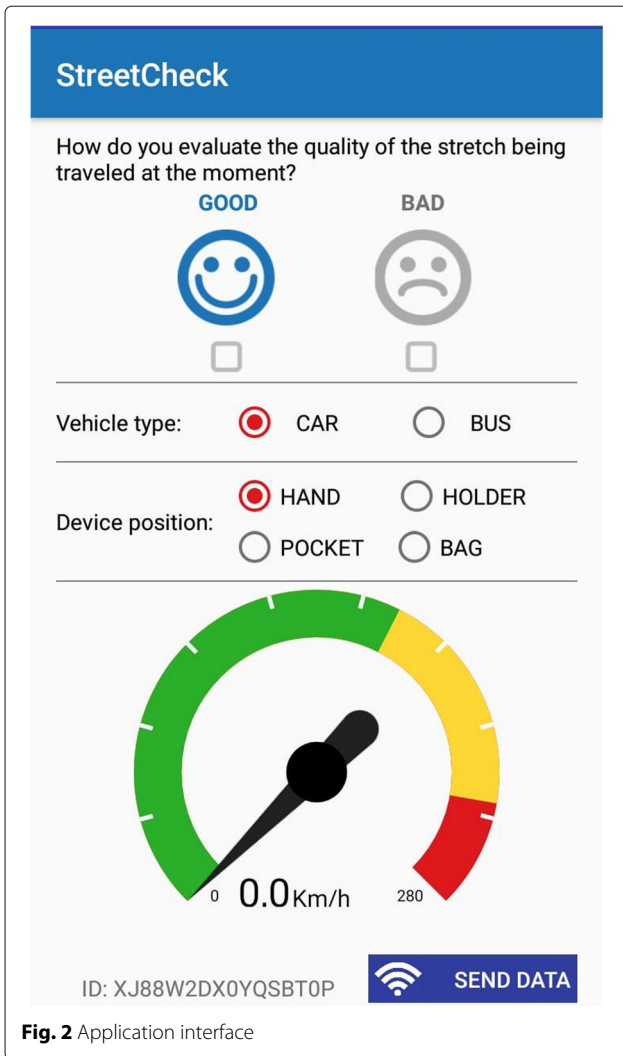
Fig. 2 Application interface

Table 2 Data gathered by the mobile device

Name	Source	Type	Meaning
Average Speed (AS)	GPS	Raw Numeric	Vehicle speed
R-Sum	Software	Numeric	Vectorial sum of x, y, z-Accelerometer axes
R-Mean	Software	Numeric	Average of R-SUM
Standard deviation (R-SD)	Software	Numeric	standard deviation of R-Mean
Latitude	GPS	Numeric	Local latitude
Longitude	GPS	Numeric	Local longitude
Label	User	Character	User opinion
Ground type	User	Character	ground type
Position	User	Character	Device positioning inside the car
Vehicle	User	Character	Vehicle type
Timestamp	Software	Numeric	Timestamp

The purpose of each feature is to provide a distinct vision of each aspect of the samples, improving the accuracy of the classification algorithms in the prediction process. The raw reading of devices' accelerometers, containing a 3-axis vector $\{x, y, z\}$, provides poor information on asphalt quality when using it as an isolated feature. For this reason, we have proposed a set of features based on the transformation of the 3-axis readings of the accelerometer. Table 3 summarizes these features.

Due to the individual interpretation of road surface conditions, different samples overlapping the same stretches of streets may be labeled with conflicting ratings.In fact, classification algorithms may interpret as noise samples with similar values but from different classes. To understand and overcome this issue, we have proposed two filters in the dataset: 1) *Consensus*; and 2) *Exclusion*. We have analyzed the accuracy of the supervised learning algorithms for each filter. These filters are explained below:

Consensus: First, the labeled dataset is clustered. Next, the filter counts how many samples belong to each
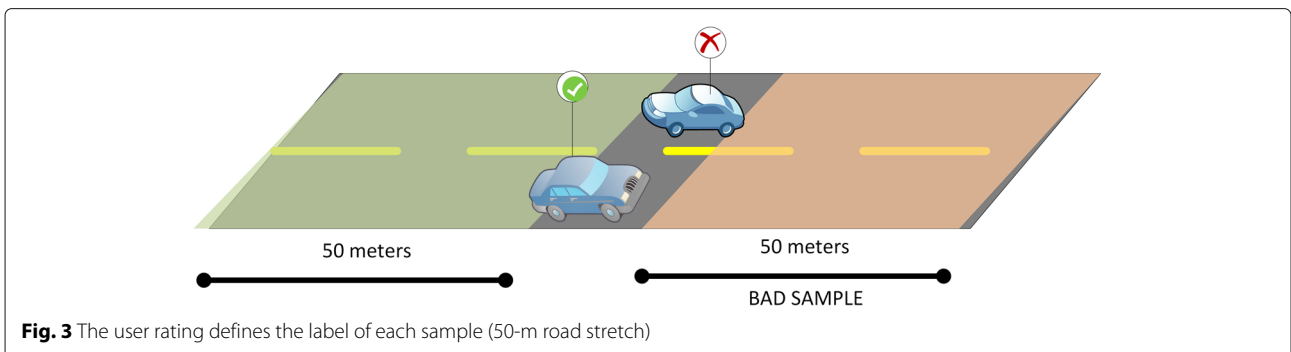
In these classifiers, samples are clustered according to the similarities found in their characteristics. The time complexity of these algorithms is $O(v * nlog(n))$ for Random Forest, $O(vn + kv)$ for the KNN, $O(n^2)$ for J48, where n is the number of records, v is the number of features, k the number of neighbors per cluster. The time complexity of SVM depends on the number of supporting vectors.

Fig. 3 The user rating defines the label of each sample (50-m road stretch)

Table 3 Set of features based on the raw data

Feature	Generation moment
Resulting Vector	Instantly
Speed	Instantly
Sum	Preprocessing, on the application
Standard deviation	Preprocessing, on the application
Softness	Preprocessing, on the application
Distortion	Post-processing, on the server
Quartiles	Post-processing, on the server

cluster. Finally, the predominant label is assigned to other samples of the same cluster. For instance, if a cluster has ten samples labeled as GOOD and five samples labeled as BAD then, the label GOOD is considered a predominant one, and the filter assigns as GOOD those five samples assigned as BAD by the users.

Exclusion: The initial process is similar to the one presented above.. However, instead of changing labels, this filter removes the non-predominant samples from the cluster to create a new training dataset. By Using the previous example, this filter removes the five samples assigned as BAD by the user from the training dataset.

The set of features were extracted in three moments: (i) instantly, after each accelerometer vector reading; (ii) preprocessing, *Streetcheck* extracts the feature every 50 meters; and, finally, (iii) post-processing, features extracted after the data transmission. Each feature is detailed below.

3.3.1 Resulting vector - R
The resulting vector R is composed of the vectorial sum of three-axis $\{x, y, z\}$ provided by the accelerometer.

Figure 4a shows the magnitude of the resulting vector compared to the three axes of the accelerometer. Figure 4b illustrates the differences between the resulting vector in samples labeled as "GOOD" and as "BAD". The vectorial sum can overcome the issue caused by the positioning of the device inside a car. That is, the resulting vector R makes Streetcheck work independently from the position of the devices. Therefore, this feature represents a basis to extract new features.

3.3.2 Average speed - AS
Vehicle speed reflects on the accelerometer readings. Therefore, the speed must be considered to calculate features. *Streetcheck* gathers vehicle speed every second and saves the average speed - *AS* - for each sample (50 meters).

3.3.3 Average resulting vector - R-Mean
The average of resulting vectors (R-Mean) for a sample reflects road conditions in a summarized fashion. Figure 5a illustrates the average resulting vectors for some labeled samples.

3.3.4 Standard deviation of the resulting vector - R-SD
The standard deviation $(R - SD)$ indicates how dispersed the sample is in relation to the average. Thus, peaks in the standard deviation of the resulting vector (R) represents the roughness of the road. Figure 5b illustrates this feature for some labeled samples. Peaks in the samples labeled as GOOD in the figure may be caused by potholes, junctions, or abrupt road anomalies.

3.3.5 Softness - Soft
The feature defined as softness (*Soft*) aims to normalize the standard deviation of the resulting vector and the average speed of the vehicle on the interval $[0, 1]$. Softness is defined by Eq. 1, where *AS* and *R-SD* are the average speed
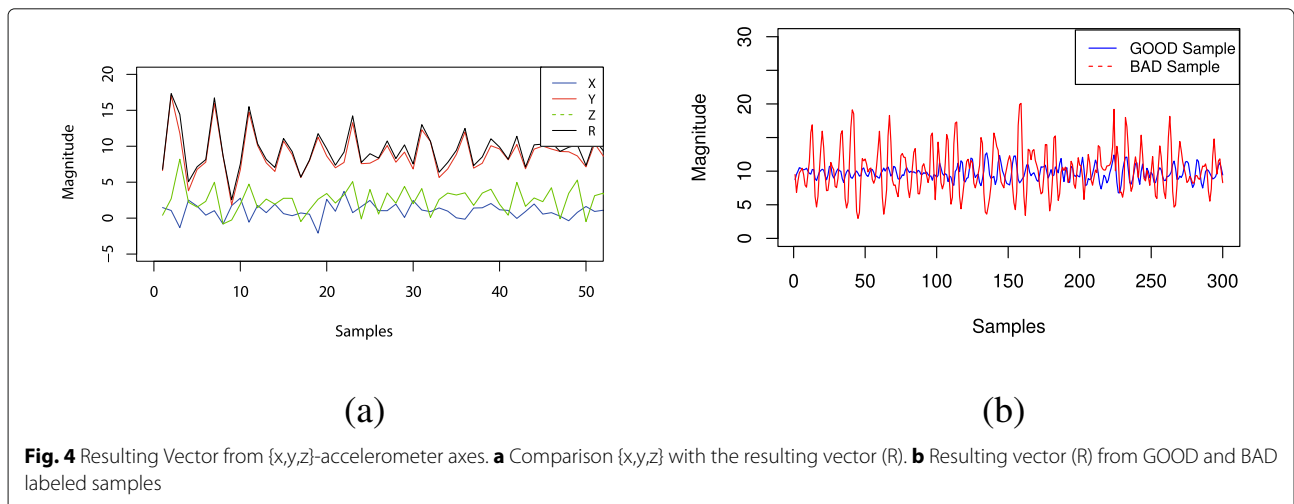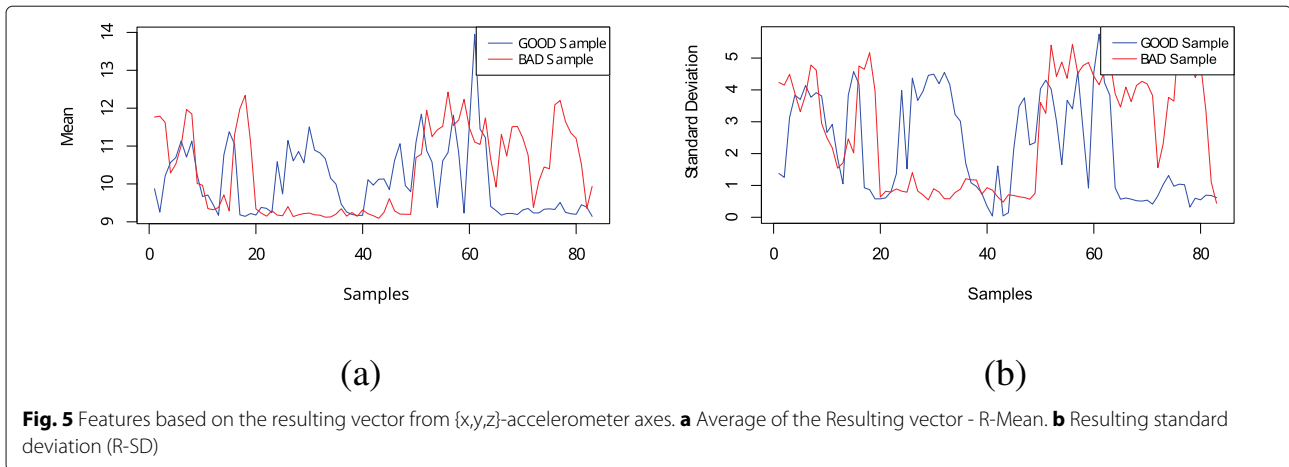

(a)   (b)

Fig. 4 Resulting Vector from {x,y,z}-accelerometer axes. **a** Comparison {x,y,z} with the resulting vector (R). **b** Resulting vector (R) from GOOD and BAD labeled samples

Fig. 5 Features based on the resulting vector from {x,y,z}-accelerometer axes. **a** Average of the Resulting vector - R-Mean. **b** Resulting standard deviation (R-SD)

and average standard deviation of the resulting vector in a sample, respectively.

$$Soft = \frac{(1 + R - SD)}{AS} \tag{1}$$

As shown in Fig. 6, *softness* closer to 1 means a regular sample, while a *softness* closer to 0 means anomalies on the surface of the road.

3.3.6 Average distortion - Dis
Average distortion (*Dis*) considers each triangle formed by the accelerometer's axes $\{x, y, z\}$ as a plane. The distortion calculates the angle between subsequent planes and the sum of these angles in a sample is divided by 360.

3.3.7 Quartiles - QX1, QX3, QY1, QY3, QZ1, QZ3
These features allow comparing the quartiles for each sample. The first and the third quartile have been analyzed to compare the lowest and highest values of the dataset.

Equations 2 and 3 calculate the first ($Q1_E$) and third ($Q3_E$) quartile for each axis $E \in \{x, y, z\}$, respectively. Therefore, each axis represents an independent feature.

Figure 7a, b and c illustrate the magnitude of the samples for {x,y,z}-axis, respectively. In these samples, categories "GOOD" and "BAD" differ in the axes in the third ($Q3$) x and first ($Q1$) quartiles of the y-axis, while the z-axis does not show great differences. The combination of these features may increase the accuracy rate of the supervised learning algorithms.
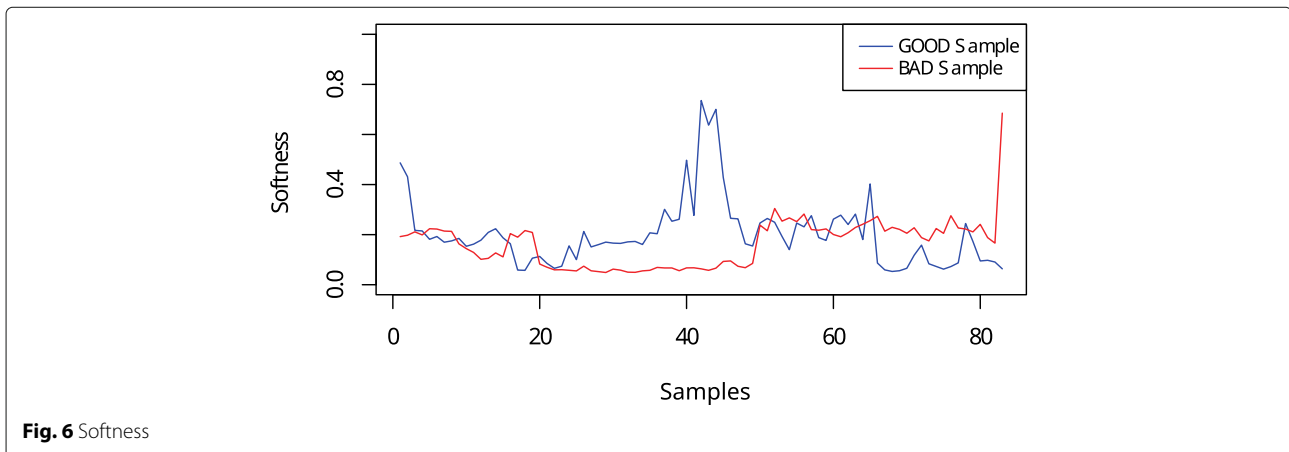
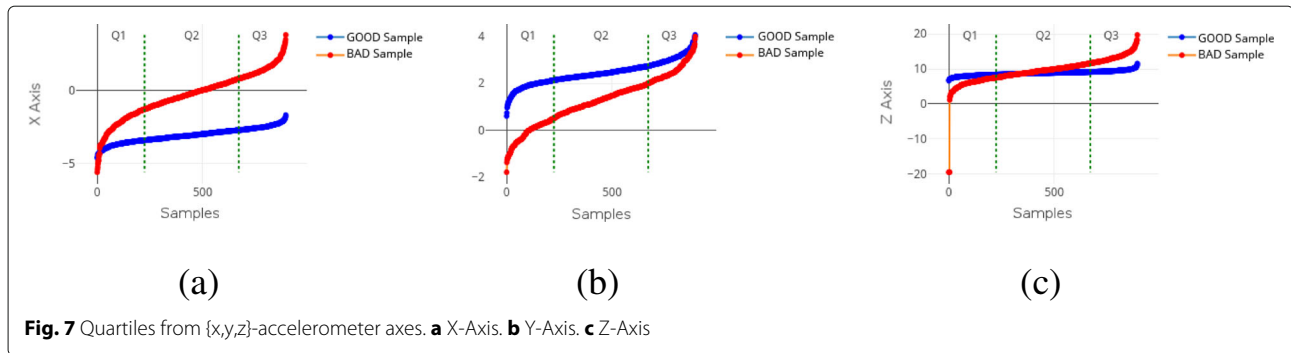$$QE1 = \frac{\sum_1^{\frac{n}{4}} Ei}{\frac{n}{4}} \tag{2}$$

$$QE3 = \frac{\sum_{(\frac{n}{4}*3)}^{n} Ei}{\frac{n}{4}} \tag{3}$$

4 Classifying road surface quality
4.1 Methodology
To validate *Streetcheck* framework and the proposed set of features, the tests were divided into two phases: i)

Fig. 6 Softness

Fig. 7 Quartiles from {x,y,z}-accelerometer axes. **a** X-Axis. **b** Y-Axis. **c** Z-Axis

data collection and, ii) data classification. We analyzed the relevance and the accuracy of the proposed features in the supervised learning algorithms Random Forest, SVM, KNN e J48.

4.1.1 Data collection
Twenty (20) volunteers installed the Streetcheck application on their personal mobile devices. Whenever possible, these volunteers rated the stretches of road they were traveling in, either by car or by bus. The Data collection phase lasted three weeks and data were gathered from different car and bus models. From the total of samples gathered, 13.74% were labeled as GOOD or BAD.

These volunteers were trained to rate their impressions about the road surface conditions when they were in the role of passengers. In the role of drivers, we asked them to let the application running in their devices, indicating the position of the device inside the car. We emphasized for all volunteers to rate the road quality only when they were passengers in the vehicle.

4.1.2 Data classification
At the second stage, we analyzed the number of clusters which better fit our dataset, the relevance of each feature, and the hit rate in the supervised algorithms: Random Forest, SVM, KNN e J48.

Our first test aimed to quantify how many representative groups exist in the dataset. We applied the *Canopy* algorithm [27] to determine the silhouette curve [28] of the dataset. As a result, the Canopy algorithm gives the number of categories expressed in the dataset.

Besides that, the presence of irrelevant or redundant features can affect the performance of classifying algorithms. The problem of selecting features can be defined as searching for M relevant characteristics from a set of N attributes, such that $M \leq N$ [29].

To define the relevance of each feature, presented in Section 3.3, we used the *Relief* algorithm [30]. *Relief* calculates a score ($\delta(X)$) for each feature X in binary class datasets. Higher values of $\delta(X)$ mean a higher relevance of the feature. As a result, we ranked the most relevant features.

We compared the performance of the supervised learning algorithms using the original training dataset and two filtered datasets. Streetcheck uses the *GNU WEKA*[5] library to implement its classification system.

Fig. 8 Distribution of samples per speed. **a** unlabeled dataset. **b** labeled dataset
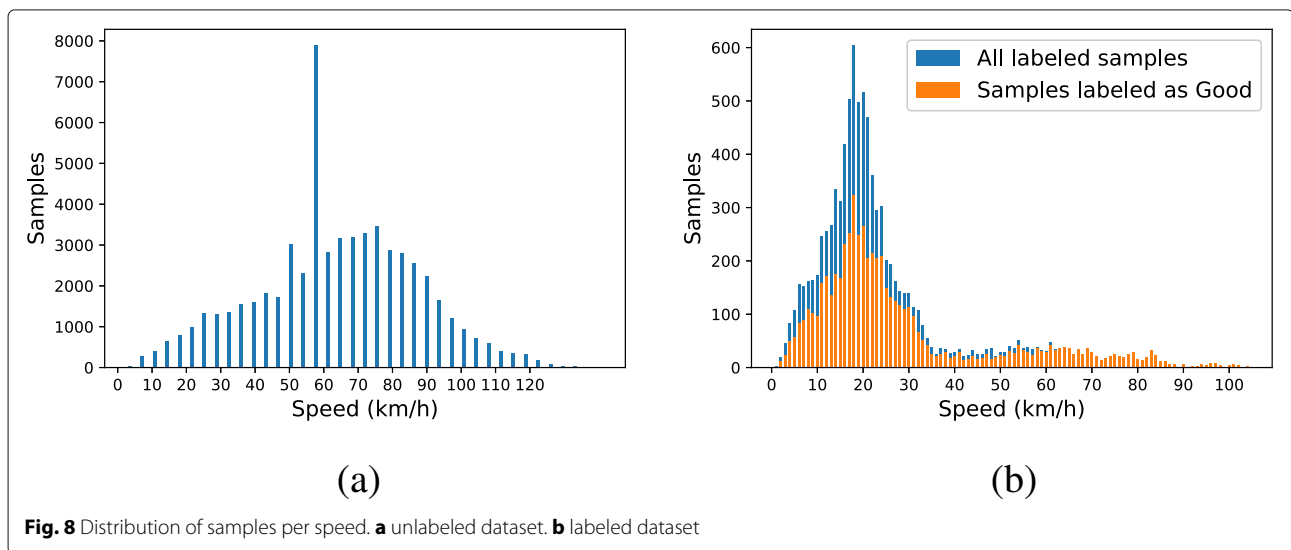
5 Dataset analysis and road conditions classification

5.1 Dataset characterization

Our volunteers traveled, as drivers or passengers, 1,218.34 km on urban roads of Minas Gerais(Brazil). Volunteers gathered 2.4GB of raw data from their smartphones. As it has been previously mentioned, raw data accelerometers readings were summarized into 50-meter samples, totalizing 68,264 samples. The training dataset is composed of 9,832 (13.74%) samples labeled, where 53.72% of these samples were labeled as GOOD and 36.28% labeled as BAD. Few samples (0.6%), covering the same stretch of the street, were labeled with different ratings, which means that two or more users provided distinct ratings about the same stretches of streets.

We analyzed the dataset attempting to answer two questions: 1) What is the influence of speed in the proposed features? 2) Which features present differences between the GOOD and BAD labeled samples?

To answer the above questions, we studied the distribution of each feature considering the speed of the vehicle and the rating. Figure 8 shows the distribution of the unlabeled dataset (Fig. 8a) and the labeled dataset (Fig. 8b). As shown in Fig. 8a, the vehicle speed follows a normal distribution, where the majority of the samples were collected under 70km/h. In the labeled dataset, users rated as "BAD" only when the vehicle was under 60km/h, as shown in Fig. 8a. These results show that if a vehicle is at a higher speed, it may mean that the road pavement is under good/satisfactory conditions.

The speed of the vehicle directly influences the standard deviation of the resulting vector (R-SD) and the softness. Figure 9 depicts the dispersion of these features based on the speed[6]. Figure 9a shows $RS-D$ presents greater values when users rate the road as being under a bad or /unsat-

isfactory conditions. The same occurs with the softness feature, Fig. 9b. The Softness fits the exponential model, Eq. 4, with $\{a, b, c\} = \{0.812, 0.061, 0.101\}$. These results indicate that R-SD and softness combined with the average speed represent good inputs for supervised learning algorithms.

$$y = ae^{(-bx)}c \tag{4}$$

Figure 10 shows the cumulative distribution function (CDF) for all features grouped by the label. R-SD, R-Mean, and softness distinguish samples labeled as GOOD from those labeled as BAD in a clear fashion, as shown in Fig. 10a, b, and c, respectively. We discuss the relevance of each feature for supervised learning algorithms in Section 5.3.

5.2 Number of categories

We used the *Canopy* algorithm and the silhouette technique to identify clusters in the dataset based on our proposed features. Initially, the *Canopy* algorithm identified and provided an insight into the spreading of the data. Figure 11 illustrates data spreading when considering speed (axis x) and softness (axis y) as features. In such a case, 75% of the dataset fits in two clusters. Furthermore, we observed that increasing speed decreases the softness. It is important to note that Fig. 11 illustrates the clustering considering only two features.

In order to verify the number of clusters that represent the dataset when all features are considered, we used the silhouette technique. The silhouette technique evaluates how data are adjusted to the clusters. Table 4 shows the silhouette factor for each number of clusters from unclassified data.
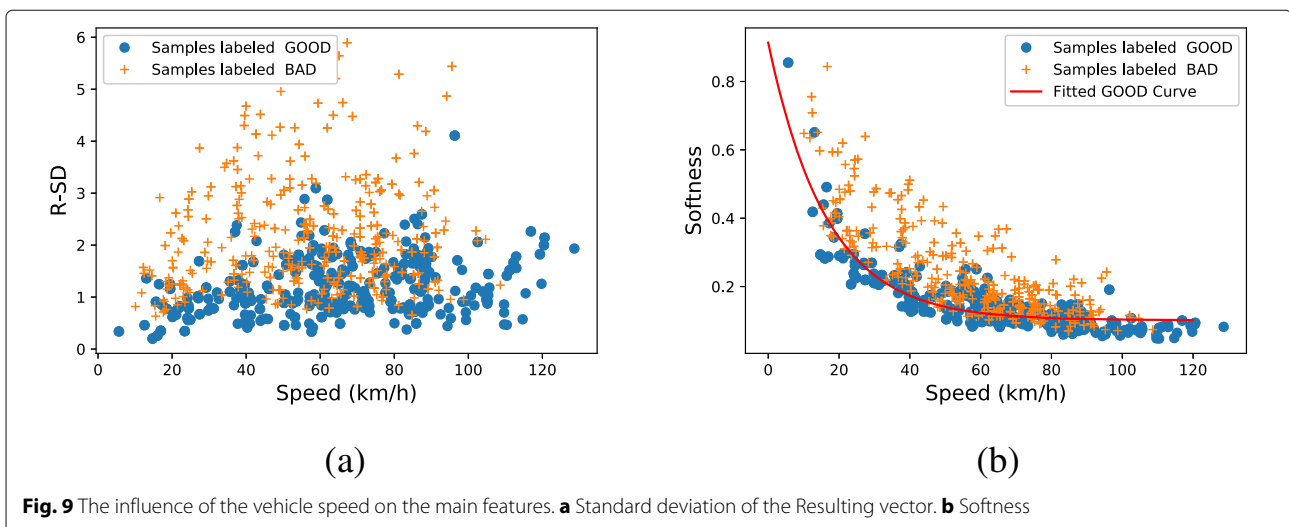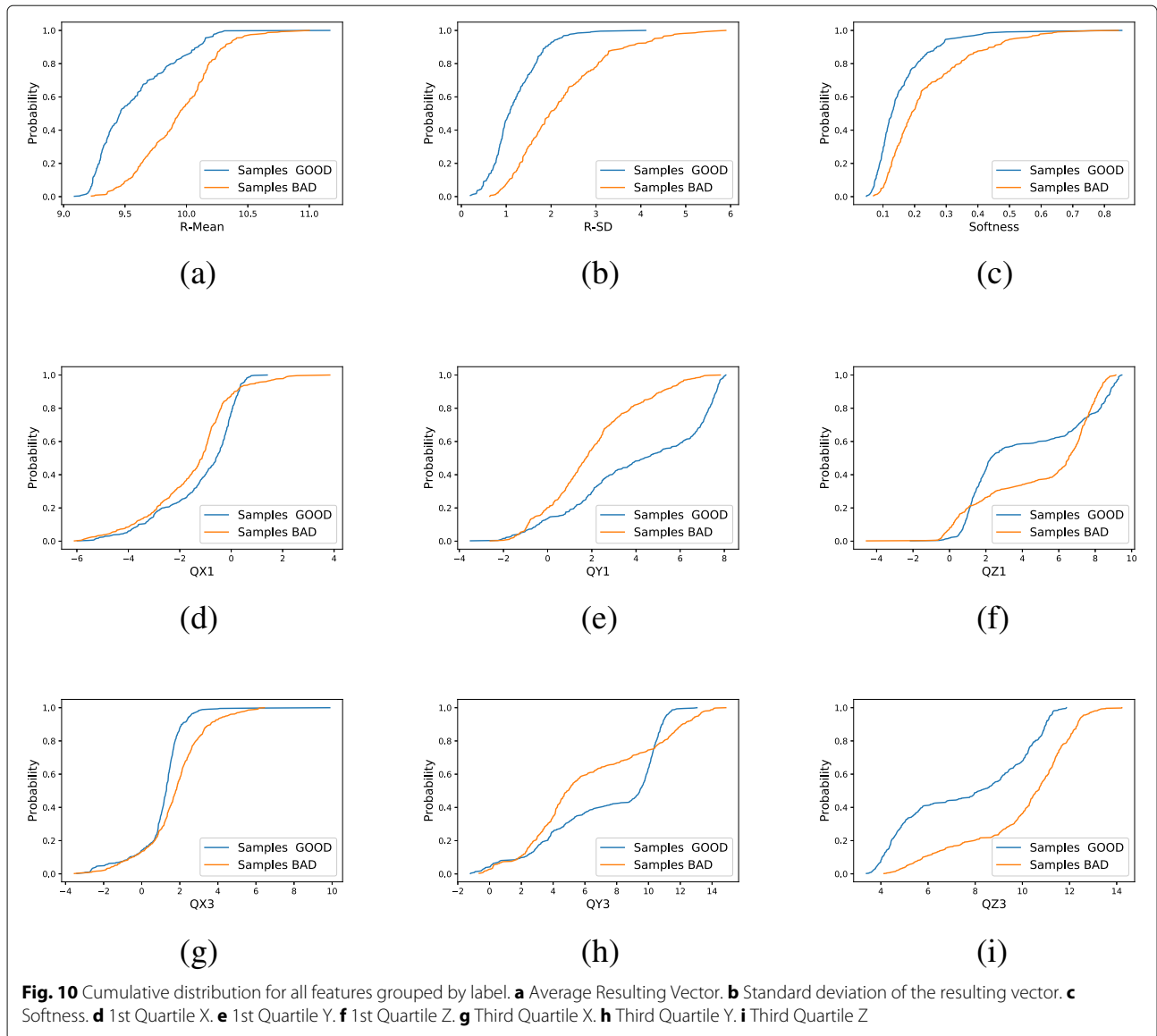
Fig. 9 The influence of the vehicle speed on the main features. **a** Standard deviation of the Resulting vector. **b** Softness

Fig. 10 Cumulative distribution for all features grouped by label. **a** Average Resulting Vector. **b** Standard deviation of the resulting vector. **c** Softness. **d** 1st Quartile X. **e** 1st Quartile Y. **f** 1st Quartile Z. **g** Third Quartile X. **h** Third Quartile Y. **i** Third Quartile Z

It can be observed that, as the number of categories increases, the silhouette factor decreases. Therefore, two classes of data are suitable to classify road surface quality. For this reason, *Streetcheck* classifies road surface in a binary mode ("GOOD" or "BAD").

5.3 Accuracy evaluation

In this section, we evaluate the relevance of the mapped features and the accuracy of road surface quality classification for the KNN, Random Forest, J48, and SVM classifiers [31]. The training dataset contains 9.382 samples labeled ("GOOD" or "BAD"). We evaluate the performance using the original dataset and the dataset modified by the *consensus* and *exclusion* filters, as detailed in Section 3.3. The results with both filters were similar. Since the exclusion filter avoids

noise in the dataset, we show the results using this filter.

The Relief algorithm scores the relevance of each feature according to their ability to differentiate the samples [32]. The average speed, softness, third and first quartile, and the standard deviation are the features with a higher relevance, according to the Relief algorithm. We argue that the combination of features with lower relevance can increase the accuracy of the classifiers. Table 5 presents the relevance of the eleven features considered in this work.

To support our statement, we tested each classifier with a subset of features. Figure 12 shows the classifiers' accuracy to the number of features, using the original training dataset. In this figure, the x-axis represents the number of features considered by the classifier. In each iteration, tests
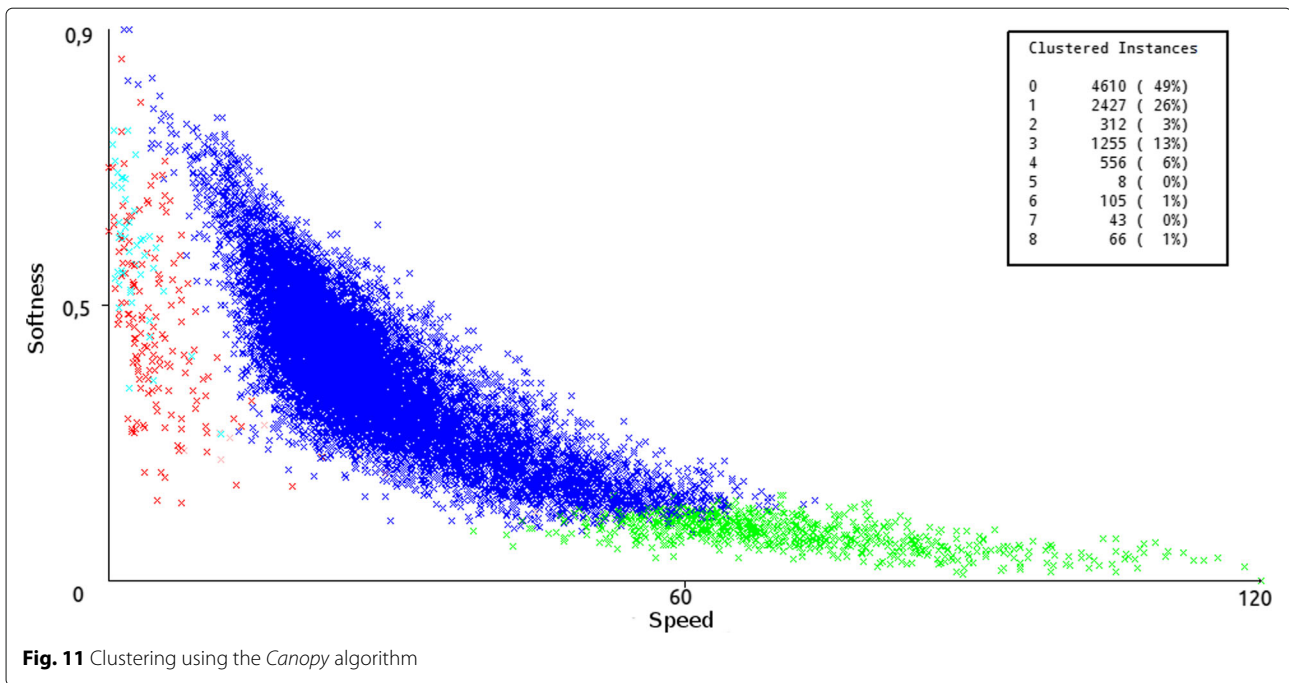
Fig. 11 Clustering using the *Canopy* algorithm

were repeated removing the less relevant feature from the classifiers, as shown in Table 5.

Classifiers reached up to 70% of accuracy when using only average speed feature. Furthermore, the accuracy of all classifiers remained stable after using eight or more features.

While SVM, KNN, and J48 classifiers performed similarly for any number of features, Random Forest obtained the highest accuracy out of all tests. Random Forest uses different decision trees to classify the samples. In such case, these multi-trees distinguished the samples more effectively.

As already mentioned, users may attribute conflicting opinions about similar stretches of street quality, which makes it harder for classifiers to distinguish between similar samples. To overcome this issue, we applied a filter to normalize and another to exclude conflicting labels. Consensus filter modifies labels of the samples that are the minority in their cluster. Meanwhile, the exclusion filter excludes these samples.

The accuracy of the classifiers using these three versions of training dataset and all features is shown in Fig. 13. Furthermore, Table 6 details the numeric results. Random Forest accuracy and SVM improved their accuracy by 9.3% and 29.87%, respectively, when comparing

their performance in the original dataset and the exclusion filtered dataset. The SVM classifier had the major improvement while filtering the original dataset, which demonstrates our approach can provide better results on classifying road surface quality. By removing conflicting samples from the training dataset, it reduces the noise for classifiers.

In both filters, labels were ranked, and the label with the highest rank defined which samples have their labels modified or excluded from the training dataset. These results show how human labeling create noise, which is harder to identify by classifiers algorithms.

5.4 Visualization

In Participatory sensing networks, individual collaboration becomes collective information. After sensing and transmitting data through the *Streetcheck*, a user may have the global view of road surface quality, already sensed by other users. Figure 14 shows the classification of one user on the map, where blue and red dots represent the stretches of streets classified as "GOOD" and "BAD", respectively.

However, data collectively gathered by users provide a global view of classified roads. Figure 15 presents the participatory sensing result: a map with all data classified by Streetcheck.

Information on the quality of a given road surface is (now) made available to users who have not taken a certain route. This information can be taken into account so that the user can make the decision on which route could be taken.

Table 4 Sillhouete factor for clusters

Clusters	2	3	4	5	6	8	10	15	20
Silhouette	0.75	0.55	0.48	0.44	0.37	0.36	0.36	0.33	0.32

Table 5 Relevance of features

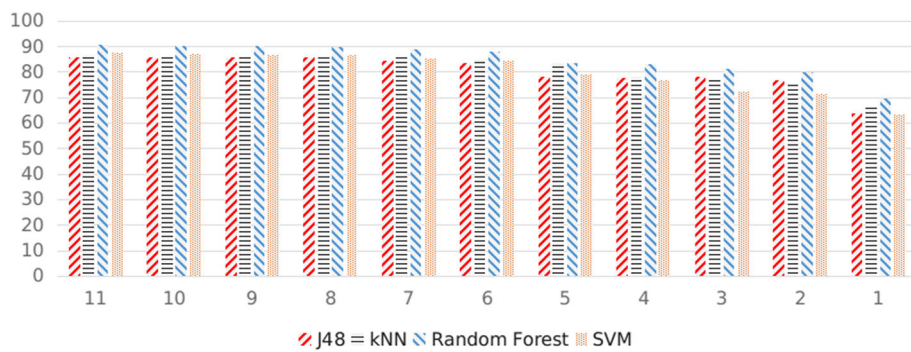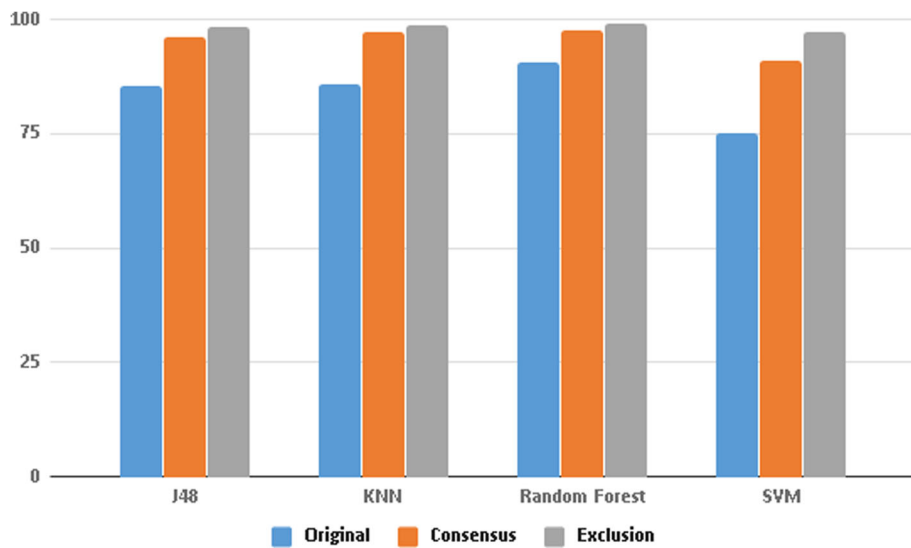| Feature	AS	Soft	QX3	QX1	R-SD	QZ1	QY1	DIS	QY3	DIS	QY3
Score ($\times 10^{-2}$)	3.079	1.328	1.228	1.202	1.032	0.949	0.916	0.882	0.835	0.825	0.795

Fig. 12 Algorithms' accuracy in relation to the number of features

Fig. 13 Accuracy of the classifiers using the original and the filtered datasets

Table 6 Accuracy of classifiers for each dataset and using all features

Algorithm	Original dataset	Consensus filter	Exclusion filter
Random Forest	90.64%	97.84%	99.11%
Knn 5 neighbors	86.04%	97.3%	98.83%
J48	85.67%	96.14%	98.4%
SVM	74.97%	91.16%	97.36%

6 Conclusions

This work presents *Streetcheck*: a framework for sensing, gathering, and classifying road surface quality based on users opinions. The framework is composed of an application for mobile devices and a classification system. The mobile application gathers accelerometer and geolocation sensor information, as well as users' opinions about stretches of streets they have passed on. The classification system filters and extracts a set of features from the aggregated data. Then, it creates a model using supervised learning classifiers to classify unlabeled data.

Since Streetchek relies on crowdsensing, users can rate with conflicting opinions about road conditions. For instance, samples of 50 meters of stretches of roads with similar accelerometer reading values were classified as GOOD and BAD by different users. Indeed, users may have different perceptions of the quality even for the same stretch of street.

This work attempts to overcome the issue mentioned above by extracting eleven features from the raw data, and also by applying filters into the dataset to remove inconsistent classification. We evaluate the accuracy with aid of five supervised learning classifiers. Our evaluation, with 20 volunteers using Streetcheck, shows that the speed of the vehicle, standard deviation, and first and third quartile of raw accelerometer readings for each sample are the most critical features to classify road segments. Therefore, we proposed a feature, called Softness, to normalize these values.

Our approach to filtering the original dataset improved the accuracy for all classifiers. The filters improved the accuracy from 9.4% (Random Forest) to 29.87% (SVM). In all cases the accuracy was higher than 97%. These filters normalized samples in the same clusters in the training dataset.

As future work, we first intend to improve the graphical interface to encourage more users gathering and labeling road surface quality. Next, we plan to provide an API to make the classification available for other services. In such case, route decision systems may use our classification results. Finally, we also aim other applications for *Streetcheck*, such as measure users' comfort in public transportation, identify and classify different pavement types. Although being out of the scope of this article, privacy, security, and the correctness of sensed raw data are still open issues that require further investigation.
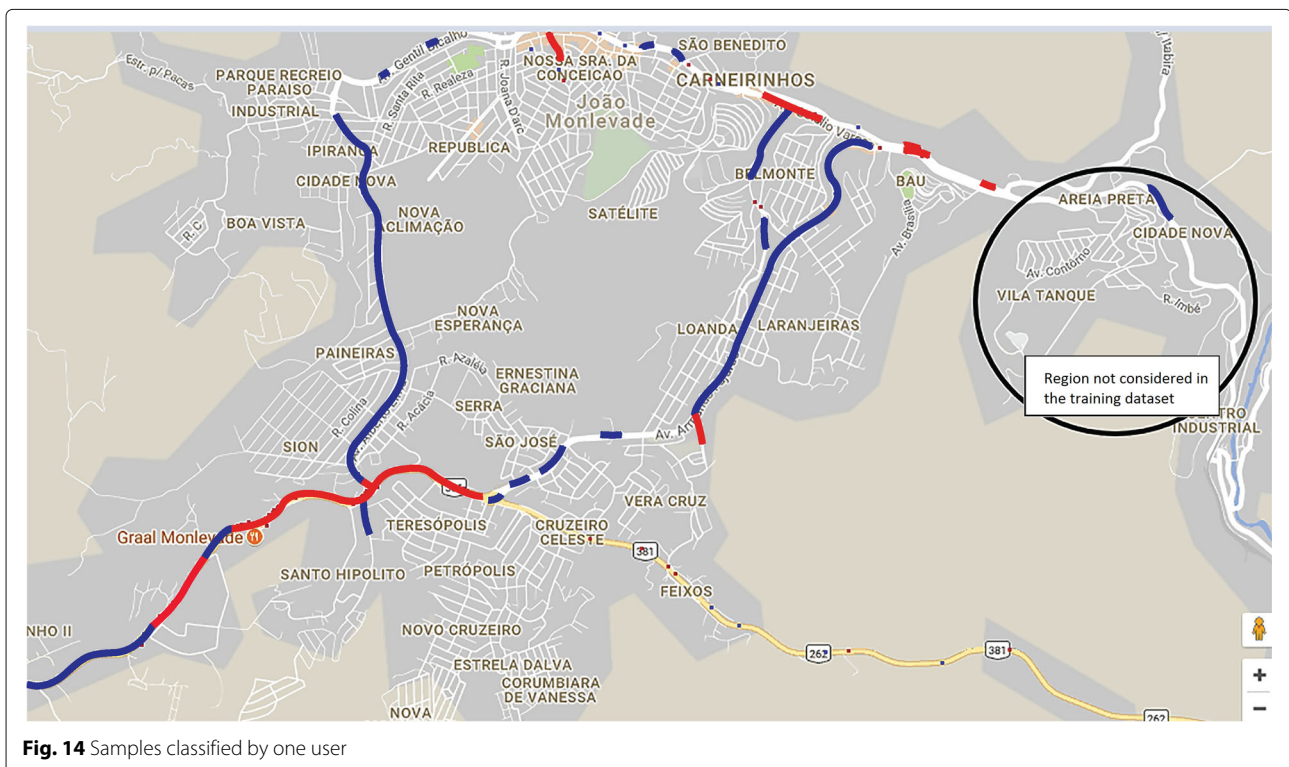
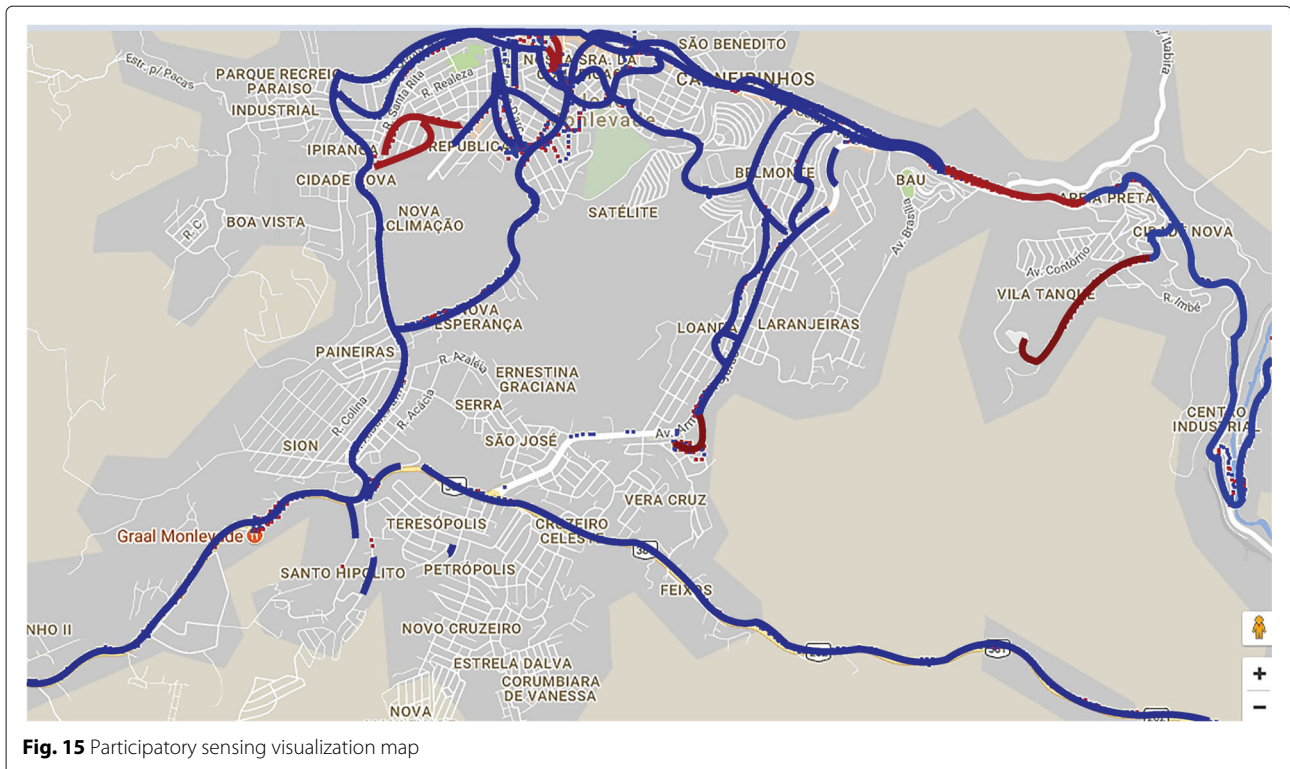

Fig. 14 Samples classified by one user

Fig. 15 Participatory sensing visualization map

Endnotes

[1] http://www.waze.com

[2] http://transito.maplink.global

[3] The standard STPM 222-5 defines 100-m stretches as samples to evaluate the IRI [33]. Since we evaluate accelerometer data without specific calibration, we used 50-meter stretches of road as a sample to evaluate the roughness of the road.

[4] Streetcheck is available on Google PlayStore.

[5] https://www.cs.waikato.ac.nz/ml/weka/

[6] For the sake of clarity, we omitted the results of features without clear difference among GOOD and BAD labels grouped by speed.

Authors' contributions
DEN made substantial contributions to the conception, design and implementation of the proposed solution. VFSM made substantial contributions to the conception and design of the proposed solution, to designing the experiments conducted, and was a major contributor in writing the manuscript. All authors read and approved the final manuscript.

Availability of data and materials
The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests
The authors declare that they have no competing interests.

Author details
[1] Department of Computer and Systems, Federal university of Ouro Preto, João Monlevade, Brazil. [2] Department of Informatics, Universidade Federal do Espírito Santo, Vitória, Brazil. [3] DEN and AF are enrolled in the Department of Computer Science, Federal University of Minas Gerais, where the research was partially conducted., Belo Horizonte, Brazil.

References
1. Wolff MGdC, Caldas MAF. A model for the evaluation of brazilian road transport: A sustainable perspective. J Adv Transp. 2018;2018:1–12.
2. Burke J, Estrin D, Hansen M, Parker A, Ramanathan N, Reddy S, Srivastava M. Participatory sensing. In: Workshop on World-Sensor-Web (WSW) at SenSys. Boulder: ACM; 2006. p. 117–34.
3. Silva TH, de Melo POV, Neto JB, IJT A, Ribeiro CSdS, Mota VF, da Cunha FD, Ferreira AP, Almeida JM, Loureiro AA. Users in the Urban Sensing Process: Challenges and Research Opportunities. In: Pervasive Computing: Next Generation Platforms for Intelligent Data Collection. Intelligent Data-Centric Systems: Sensor Collected Intelligence. New York: Elsevier; 2016.
4. Conti M, Passarella A. The internet of people: A human and data-centric paradigm for the next generation internet. Comput Commun. 2018;131: 51–65.
5. Ndoye M, Barker AM, Krogmeier JV, Bullock DM. A recursive multiscale correlation-averaging algorithm for an automated distributed road-condition-monitoring system. IEEE Trans Intell Transp Syst. 2011;12(3):795–808.
6. Kumar R, Mukherjee A, Singh V. Community sensor network for monitoring road roughness using smartphones. J Comput Civil Eng. 2016;31(3):04016059.

7. Mota VF, Silva TH, Macedo DF, Ghamri-Doudane Y, Nogueira JM. Towards scalable mobile crowdsensing through device-to-device communication. J Netw Comput Appl. 2018;122:99–106.

8. Alpaydin E. Introduction to Machine Learning. Cambridge: MIT Press; 2014.

9. Boutaba R, Salahuddin MA, Limam N, Ayoubi S, Shahriar N, Estrada-Solano F, Caicedo OM. A comprehensive survey on machine learning for networking: evolution, applications and research opportunities. J Internet Serv Appl. 2018;9(1):16.

10. González A, O'brien EJ, Li Y-Y, Cashell K. The use of vehicle acceleration measurements to estimate road roughness. Veh Syst Dyn. 2008;46(6): 483–99.

11. Mednis A, Strazdins G, Zviedris R, Kanonirs G, Selavo L. Real time pothole detection using Android smartphones with accelerometers. In: Distributed Computing in Sensor Systems and Workshops (DCOSS). Barcelona: IEEE; 2011. p. 1–6.

12. Ng JR, Wong JS, Goh VT, Yap WJ, Yap TTV, Ng H. Identification of road surface conditions using iot sensors and machine learning, vol. 481. Singapore: Springer; 2019. p. 259–68.

13. Eriksson J, Girod L, Hull B, Newton R, Madden S, Balakrishnan H. The pothole patrol: using a mobile sensor network for road surface monitoring. In: International Conference on Mobile Systems, Applications, and Services. New York: ACM; 2008. p. 29–39.

14. Astarita V, Caruso MV, Danieli G, Festa DC, Giofrè VP, Iuele T, Vaiana R. A mobile application for road surface quality control: Uniqualroad. Procedia - Soc Behav Sci. 2012;54:1135–44.

15. Alessandroni G, Carini A, Lattanzi E, Freschi V, Bogliolo A. A Study on the Influence of Speed on Road Roughness Sensing: The SmartRoadSense Case. Sensors. 2017;17(2):305.

16. Lima LC, Amorim VJP, Pereira IM, Ribeiro FN, Oliveira RAR. Using Crowdsourcing Techniques and Mobile Devices for Asphaltic Pavement Quality Recognition. In: VI Brazilian Symposium on Computing Systems Engineering (SBESC). Porto Alegre: SBC; 2016. p. 144–9.

17. González LC, Moreno R, Escalante HJ, Martínez F, Carlos MR. Learning roadway surface disruption patterns using the bag of words representation. IEEE Trans Intell Transp Syst. 2017;PP(99):1–13.

18. Quintana M, Torres J, Menéndez JM. A simplified computer vision system for road surface inspection and maintenance. IEEE Trans Intell Trans Syst. 2016;17(3):608–19.

19. Douangphachanh V, Oneyama H. A study on the use of smartphones under realistic settings to estimate road roughness condition. EURASIP J Wirel Commun Netw. 2014;2014(1):114.

20. Piao B, Aihara K, Kinoshita A, Takasu A, Adachi J. Estimating road surface condition using crowdsourcing. In: Information Search, Integration, and Personlization. Cham: Springer; 2017. p. 66–81.

21. Singh G, Bansal D, Sofat S, Aggarwal N. Smart patrolling: An efficient road surface monitoring using smartphone sensors and crowdsourcing. Pervasive Mob Comput. 2017;40:71–88.

22. Chen K, Lu M, Tan G, Wu J. Crsm: Crowdsourcing based road surface monitoring. In: IEEE International Conference on High Performance Computing and Communications. Zhangjiajie: IEEE; 2013. p. 2151–8.

23. Yi C-W, Chuang Y-T, Nian C-S. Toward crowdsourcing-based road pavement monitoring by mobile sensing technologies. IEEE Trans Intell Transp Syst. 2015;16(4):1905–17.

24. Mohamed A, Fouad MMM, Elhariri E, El-Bendary N, Zawbaa HM, Tahoun M, Hassanien AE. Roadmonitor: An intelligent road surface condition monitoring system. In: Intelligent Systems. Cham: Springer; 2015. p. 377–87.

25. Doering J, Happel J, Tharmakularajah L, Krieger K-L. Road surface condition detection with planar capacitive sensors. In: Sensors and Measuring Systems ITG/GMA-Symposium. Nuremberg: VDE; 2018. p. 1–4.

26. Android. SensorEvent - Android Developers. 2019. http://developer.android.com/reference/android/hardware/SensorEvent.html. Accessed 11 June 2019.

27. McCallum A, Nigam K, Ungar LH. Efficient clustering of high-dimensional data sets with application to reference matching. In: ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD). Boston: ACM; 2000. p. 169–78.

28. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math. 1987;20:53–65.

29. Liu H, Setiono R. Feature selection and classification: A probabilistic wrapper approach. In: Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems. Fukuoka: Gordon and Breach Science Publishers; 1996. p. 419–24.

30. Kira K, Rendell LA. The feature selection problem: traditional methods and a new algorithm. In: Proceedings of the Conference on Artificial Intelligence. San Jose: AAAI Press; 1992. p. 129–34.

31. Kotsiantis SB, Zaharakis I, Pintelas P. Supervised machine learning: A review of classification techniques. Emerg Artif Intell Appl Comput Eng. 2007;160:3–24.

32. Liu H, Motoda H. Computational Methods of Feature Selection. Florida: CRC Press; 2007.

33. STPM. Roughness Profile With High Speed Profilograph. In: Standard Test Procedures Manual: STP222-5. Regina: Saskatchewan Highways and Transportation; 2003.

Publisher's Note