Learning Roadway Surface Disruption Patterns Using the Bag of Words Representation

Luis C. González, *Member, IEEE*, Ricardo Moreno, Hugo Jair Escalante, Fernando Martínez, and Manuel Ricardo Carlos

***Abstract*—Accurately classifying roadway surface disruptions (RSDs) plays a crucial role to enhance quality transportation and road safety. To this end, smartphones are becoming an ad hoc tool to collect road data, while the user is at the steering wheel. In this paper, for the first time, sensed data are represented with a novel technique inspired in the bag of words representation. New results suggest that segments of accelerometer readings play a key role to characterize different classes of events, boosting classification performance. A novel data collection process was conducted in real-life environments, where the smartphones were freely placed at five user-surveyed locations, within a fleet of cars and trucks. To the best of our knowledge, this is the largest and most heterogenous data set for RSDs, and we make it publicly available. We approach the problem of identifying RSDs as one of supervised learning, where we contrast representative classifiers, most of them not previously reported. We exhaustively evaluated the performance of all classifiers in six data sets, most of them resembling actual data sets used in similar projects. We found that in all cases, the best classifier outperforms the best results reported so far. The proposed methodology was extensively evaluated through a sensitivity analysis to determine the relevance of the parameters. Experimental results reveal that the representation technique boosts considerably the classification performance when compared with the state of the art solutions, reducing in one order of magnitude the false-positives/negatives rate and surpassing the classification accuracy for about 10% in a multiclass data set.**

***Index Terms*—Roadway surface disruptions, accelerometer, bag of words representation, machine learning.**

I. INTRODUCTION

WITH the democratization of technology, new and cheaper ways are being proposed to tackle problems that otherwise would have been thought to be highly resource dependent. Recently, there is an increasing interest to use average cars as a part of sensing systems to monitor disruptive events such as the presence of potholes, cracks, speed bumps and other anomalies in transportation infrastructure e.g., roads, streets, boulevards and highways. This tendency has become a hot topic, even being researched by big car manufacturers.[1] To this end, two approaches that have been followed are either using image processing techniques [1], [2] or embedded inertial sensors [3]. Nonetheless, both approaches require additional investment or constant inspection and maintenance. A possible alternative that has received attention from the community is to take advantage of the drivers' smartphones. Since they are equipped with a bunch of internal sensors with rich possibilities to measure environmental variables, smartphones are $ad-hoc$, affordable and capable devices for sensing the road.

In the literature, there are ongoing research projects (e.g., Pothole Patrol [4] in the city of Boston) that use smartphones' accelerometer signals to detect road anomalies, also known as Roadway Surface Disruptions (RSDs). An RSD is any permanent obstacle generated by the continuous use, weather conditions, or traffic planning decisions in the road. Although a good maintenance of roadways infrastructure is crucial worldwide, the presence of RSDs on roads, streets, boulevards and highways is ubiquitous. Their presence can cause continuous wear to vehicle suspensions and damage to tires up to fatal accidents. Note that *AAA motor club* estimated for 2014, a 6.4 billion dollars of damages caused to vehicles by potholes only in the U.S.[2]

The accelerometer sensor measures inertial forces affecting the smartphone, not surprisingly this information has been exploited in the literature to detect road anomalies when a smartphone-user is at the steering wheel. Each of the accelerometer axes corresponds to longitudinal (y-axis), vertical (z-axis) and transversal (x-axis) axes of the smartphone (see Fig. 1). When the smartphone experiences an acceleration in any of these axes, the accelerometer captures it (in m/s^2). The problem then, is to go the other way back, from sensor (accelerometer) data to inferring what forces were present and given the context, gain some knowledge from the phenomenon. It is important to note that the actual perturbation on specific axes depends on the orientation of the smartphone when an acceleration is sensed. This consideration may prove to be very helpful to know what axis would be of major importance when analyzing the impact of an anomaly, e.g., the axis that is perpendicular to the road.

Through a careful analysis of the related literature we observe that there is a good number of works that have tried to

Manuscript received February 11, 2016; revised August 23, 2016; accepted January 27, 2017. Date of publication February 21, 2017; date of current version October 30, 2017. This work was supported by CONACYT under Project CB2014-241306. The work of R. Moreno and M. R. Carlos was supported by the National Council of Science and Technology of Mexico, the agency that sponsored them via scholarships for graduate studies. The Associate Editor for this paper was Y. Chen.

L. C. González, R. Moreno, F. Martínez, and M. R. Carlos are with the Facultad de Ingeniería, Universidad Autónoma de Chihuahua, Chihuahua 31125, Mexico (e-mail: lcgonzalez@uach.mx).

H. J. Escalante is with Instituto Nacional de Astrofísica, Óptica y Electrónica, Puebla 72840, Mexico.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2017.2662483

[1] http://www.landrover.com/experiences/news/pothole-detection.html

[2] http://www.wusa9.com/story/news/nation/2014/02/24/potholes-damage-cost-us/5773501/

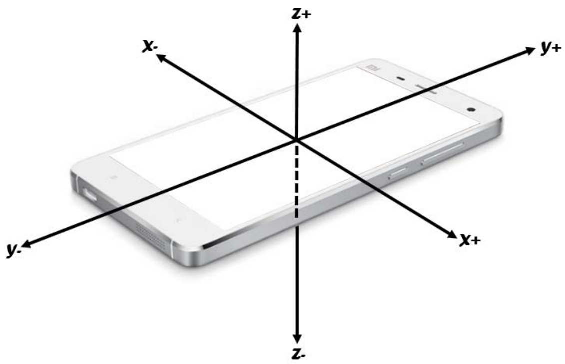


Fig. 1. Accelerometer axes with respect to a smartphone orientation.

tackle this problem. Nonetheless, there are signs that convey that maturity has not been reached yet. For example, there are no publicly available data sets to compare and contrast RSD approaches. This fundamental issue has inhibited gaining a better perspective of what has been actually accomplished and what are still the challenges for the proposed methodologies. Another sign is that previous works have missed important aspects for real-life scenarios when testing their algorithms, for example, the places where the smartphone is arranged to collect data, since it is assumed that every user will have time and interest to use a dock attached to the dash board, which goes in opposite direction of seamless and transparent technology philosophy. In addition to the lack of data and benchmarking protocols, most solutions to the classification of RSDs fail at capturing discriminative patterns in sensed data, this is evidenced in the reported high rates of False Positives and False Negatives. We hypothesize that this is due to the type of data representation adopted in such works, meaning that no significant feature vectors to accurately represent the data have been found.

Taking into consideration these issues, this paper concentrates on: *(i)* proposing a new and more heterogenous data set, *(ii)* building a better feature vector that boosts classification performance and *(iii)* to report the best accuracy results obtained by an algorithm for this problem.

More specifically, we introduce a novel data set for benchmarking RSDs classification techniques and make it publicly available. This data set was manually curated and validated to have confidence that accelerometer data represents the intended anomalies. Furthermore, we propose a novel representation for the accelerometer readings captured by the smartphones that is inspired in the Bag of Words formulation, widely used in text mining and computer vision. Its rationale is based on extracting short and representative segments of the signal to build a catalogue, by which accelerometer measurements can be represented. This representation has proved to be highly discriminative in other domains, where time series patterns were detected. We performed a sensitivity analysis on the different parameters influencing this methodology and found that a new overlapping strategy, not considered previously, boosts the performance of classification algorithms. On top of this, we report the results of several classifiers that clearly outperform current strategies regarding accuracy and False Positive/Negative Rates.

The contributions of this works are outlined as follows:

- A heterogeneous data set for RSDs that comprises 12 cars and trucks (largest number reported so far) together with the collection of data by using smartphones at new untested locations within the vehicle. This data set is being made publicly available for benchmarking.
- A novel representation inspired in the Bag of Words formulation is proposed for representing smartphone accelerometer data. The representation proved to be very useful for classifying RSDs.
- An extensive experimental study comprising seven representative classifiers, being the largest number of techniques that had been compared, from this study we report the best known classification scores for this problem.

The organization of the article is as follows. Next section reviews relevant literature. Section III explains the data acquisition process and the platform we developed to support it. Section IV presents the data representation methodology and explains how the feature vectors were generated. Section V presents the experimental results of our study, for binary and multiclass schemes, as well as a sensitivity analysis performed for the adopted representation. Finally, Section 6 outlines conclusions and suggests possible future paths for research.

II. Related Work

The problem of classifying RSDs based on data collected via mobile devices has been approached under two main paradigms. On the one hand, there are those that have tried to tackle this problem through the use of threshold heuristics. Erickson *et al.* [4] presented a system named *Pothole Patrol* that takes advantage of the presence of information beyond the z-axis, using the x-axis of the accelerometer as well. The identification of potholes reached a 92.4% of accuracy. Mohan *et al.* [5] proposed a system named Nericell that was the first work to propose a reorientation of the accelerometer axes. For bump detection they reported two scores: for low speeds, the False Negatives (FN) and False Positives (FP) rates were 37% and 14%, respectively. For high speeds, FN and FP rates were 41% and 8%, respectively. De Silva *et al.* [6] proposed a system based on filters named *BusNet* to locate potholes along the path traversed by public transportation buses. The best heuristic reported by Mednis *et al.* [7] reached a True Positive Rate of 92%. In the work of Astarita *et al.* [8] their system detected up to 90% of bump events whereas the rate of false positive events was about 35%. Douangphachanh and Onemaya [9] presented a study exploring relationships between acceleration vibration and road roughness conditions. They found that the major correlation occurs at low vehicle speeds (< 20 km/h). The system of Fazzen *et al.* [10] was able to classify on average an 85% of five different events. Sinharay *et al.* [11] based on concerns about battery consumption for smartphones motivated the need to use the accelerometer to sample at very low rates (4-6Hz). They reported a 48% FP Rate with a TP (True Positive) Rate of about 44% for pothole identification.

On the other hand, there is a less explored paradigm based on the assumption that accelerometer data can be

TABLE I
LITERATURE WORKS AND THEIR MAIN CHARACTERISTICS

Ref.	Vehicles used	What was detected	Technique	Device per vehicle
Pothole Patrol [4]	7	Potholes (92.4% accuracy)	Threshold	1 accelerometer sensor (380 Hz.)
Nericell [5]	1	Bumps (8% FP, 41% FN)	Threshold	1 accelerometer sensor (310 Hz.)
Mednis [7]	1	Potholes (92% TP)	Threshold	1 smartphone (52 Hz.)
Perttunen [12]	1	Anomalies (3% FP, 18% FN, AUC 0.97)	SVM (RBF kernel)	1 smartphone (38 Hz.)
Astarita [8]	1	Bumps (93% TP), Potholes 35% FP	Threshold	1 smartphone (5 Hz.)
Fazeen et al. [10]	1	Several classes (85.6% accuracy)	Threshold	1 smartphone (25 Hz.)
Sinharay et al. [11]	1	Bumper (85.6% TP), Potholes()	Threshold	1 smartphone (4-6 Hz.)
González et al. [14]	2	Multiclass (85.6% accuracy)	ANN and Logistic regression	1 smartphone (50 Hz.)
Seraj et al. [15]	5	Anomalies (accuracy 88.78%)	SVM (RBF kernel)	1 smartphone (47 and 93 Hz.), 1 sensor (200 Hz.)

preprocessed and modeled with Machine Learning algorithms. Perttunen *et al.* [12] presented a pattern recognition system based on Support Vector Machines to classify road segments. For bump detection they report a FP rate of 3% and FN rate of 18%. González *et. al.* [13], [14] employed two Machine Learning techniques to process accelerometer signals for events similar to the ones of Fazeen *et al.* For the experimental section they compared an Artificial Neural Network against a Logistic Regression model, obtaining an average accuracy of 83.73% over several data sets using the former classifier. Seraj *et al.* [15] made a distinction between severe and mild road anomalies, they compared several feature extraction techniques to preprocess the data. As a classifier they implemented a Support Vector Machine with a Radial Basis Function. They report an 88.78% accuracy when data is treated with wavelet transform to classify anomalous from normal road segments. As for severe and mild sub classification they achieved a 91.1% success rate.

Table I provides a summary of the related works and their distinctive characteristics. From this table and the review above, we can see that: (1) several works consider an external accelerometer (not using a smartphone), (2) most methods are based on threshold heuristics, (3) methods based on machine learning obtain results comparable to those obtained by simple thresholds, (4) the best percentage classification reaches a 92% for binary and 85% when several classes are considered.

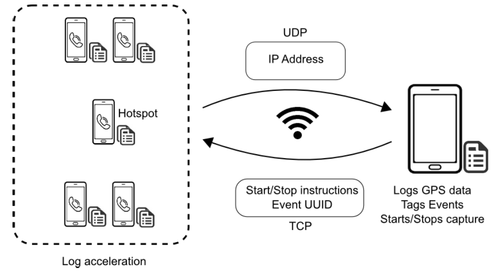


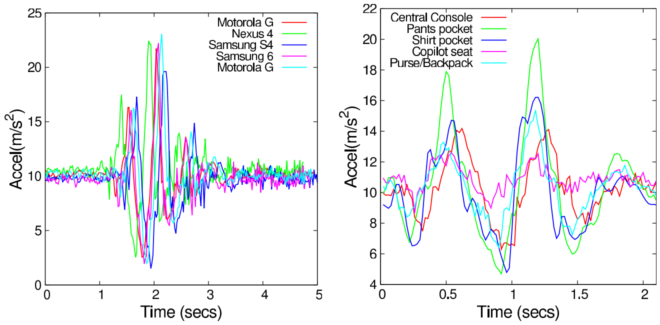
Fig. 2. Architecture of the data acquisition system.

III. DATA COLLECTION

The first contribution of this work is the introduction of a new data set for the evaluation of RSDs classification methods. As previously mentioned, no data set is publicly available for these purposes nowadays. This section describes the data collection process in detail.

We developed a platform to communicate five smartphones with a tablet that would serve as a control device to tag events. All of the data were collected in the city of Chihuahua, Mexico. The data collection process was as follows. Since we focused in the classification of RSDs, different sections of the city were identified where the events of interest, e.g., potholes, bumps, were found. From this information we constructed a catalog with longitude and latitude coordinates of these events. Given this information, the driver and his companion were aware of the exact location and nature of an event, therefore there was a planned route to follow. When a certain distance from an event, the driver would warn the copilot to start the collection process. The companion with the tablet (tagger) in his hands would choose the category of the event and the smartphones would start recording the event via the accelerometer sensors. After passing the event, the companion would finish the event collection. This process was followed to tag all the events. To make sure the event was recorded and correctly tagged, the beginning and ending of the recording was made several seconds prior and posterior to the car passing the event. To record regular road, random sections of the city were used, where no RSDs were present. To capture all these events, in total more than 130 kilometers were travelled. Some complementary statistics of the event collection process are available with our data, released online. Figure 2 shows the architecture of the data collection system.

There were several decisions that had to be made such as the model, number and locations of smartphones within the vehicle, as well as the sampling rate of the accelerometer. To conduct this study, we decided to choose a common smartphone (one of the best selling models in recent years) for the collection process. We employed five 2013 Motorola Moto G smartphones powered by a 1.2GHz processor with Android OS version 4.4.4., and tri-axial accelerometer ST Micro LIS3DH. As a reference of how multiple smartphones are compared regarding their accelerometer sensors, consider Figure 3a, where no significant differences appear among five different smartphones based on their raw accelerometer z-axis signal when placed next to each other in a car at the moment of hitting the same metal bump.



(a) (b)

Fig. 3. a) Accelerometer raw signal readings from the z-axis of five smartphones under the same event (metal bump) and placed at the same location (central console) in a vehicle. b) Example of the accelerometer z-axis raw data from five different positions in a car at the moment of capturing a pothole.

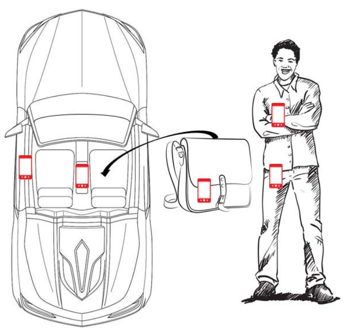


Fig. 4. Locations within the vehicle where smartphones were placed (best seen in color).

Also, we conducted a survey to determine common places where a driver would naturally place his/her smartphone while driving, the most common places were (in random order): (1) driver door, (2) center console of the vehicle, (3) driver's shirt pocket, (4) driver's pants pocket, and (5) lady purse/ backpack on the copilot seat (see Figure 4). This is the first time, to the best of our knowledge, that a work uses data collected in this way, emulating a more realistic smartphone usage. To have an idea of how accelerometer z-axis signals coming from different places at the same vehicle look when capturing the same pothole, see Figure 3b.

The literature is varied in options for the sampling rate, going from 4 to 380 Hz. After considering the energy consumption of the device, the distance granularity of using different sampling rates, and the possible lag due to the technological capabilities of the smartphone model, we decided to use a 50 Hz sampling rate. Note that the accelerometer sensor in the smartphone has, by far, a much smaller power consumption (at least two orders of magnitude in Mw) than other components such as Bluetooth or the communication components used during a call [16]. We employed 12 vehicles to perform the data collection. Several makes of vehicles were employed, and the gap between the oldest and most recent vehicle is 23 years. Table II shows the characteristics of the vehicles that were used in this study as well as the distribution of events collected per car. More than 500 events were recorded corresponding to five categories: potholes, metal bumps, asphalt bumps, worn out road, and regular road. To the best of our knowledge this is the largest and most complete data set for benchmarking RSDs classification techniques.

TABLE II
VEHICLES USED TO COLLECT THE DATA AND EVENTS TYPE PER VEHICLE

Type	Make	Model	Pothole	Speed bump	Metal bump	Worn out	Reg
Car	Nissan	Sentra 2013	3	4	6	8	7
Car	Nissan	Tsuru	1	13	33	0	0
Car	Toyota	Camry	15	1	1	17	11
Car	Chevr	Chevy	39	23	17	15	9
PickUp	Chevr	S10	1	9	2	8	3
Car	VW	Jetta	4	2	4	4	2
Car	Nissan	Sentra 2006	4	5	2	8	14
PickUp	Nissan	Frontier	9	1	6	8	11
Car	VW	Beetle	6	6	9	6	9
Car	Nissan	Altima	7	3	2	7	1
Car	Chevr	Monza	0	12	6	0	20
Car	VW	Pointer	11	21	12	19	13

IV. BAG OF WORDS REPRESENTATION

Previous works have shown that accelerometer data has been preprocessed in different ways [17], including the domains of *frequency* (e.g., Fourier/wavelet transformation) and *time* (e.g., average, standard dev, other scores, etc.). However, in most cases, raw values from accelerometer signals are used directly as features that are fed to classification models. This work proposes the use of a representation inspired in the Bag of Words (BoW) formulation for representing accelerometer data. BoW is a standard representation in text mining where documents are represented by histograms that account for the frequency of occurrence of words in the document. This representation has been extended to other domains, including, computer vision [18], speech processing [19] and even time series [20]. In these domains, a preprocessing step is performed for learning a codebook of descriptors playing the role of words in text mining. Then objects are represented by histograms accounting for the frequency of codewords in the objects. The BoW representation has proved to be very effective in all of these and other domains.

To the best of our knowledge this is the first work reporting the use of BoW for accelerometer data. In the following, we briefly review related works. For time series processing, Lin *et al* [21] proposed a technique based on *Bag of Words* that they named *bag of patterns*. The idea is to segment the complete time series and represent each segment with a *word* within a dictionary. For this purpose they used for data representation the Symbolic Aggregate approXimation (SAX). Motivated by this idea, Wang *et al.* [20], replaced SAX by the Discrete Wavelet Transform (DWT), thus reducing the dimensionality of the bag of patterns representation. These proposals have shown an enormous capacity of effectively representing complete time series, which motivated us for proposing a similar representation for accelerometer data. This fact was also supported by previous comparisons against Fourier/wavelet transform, Dynamic Time Warping and

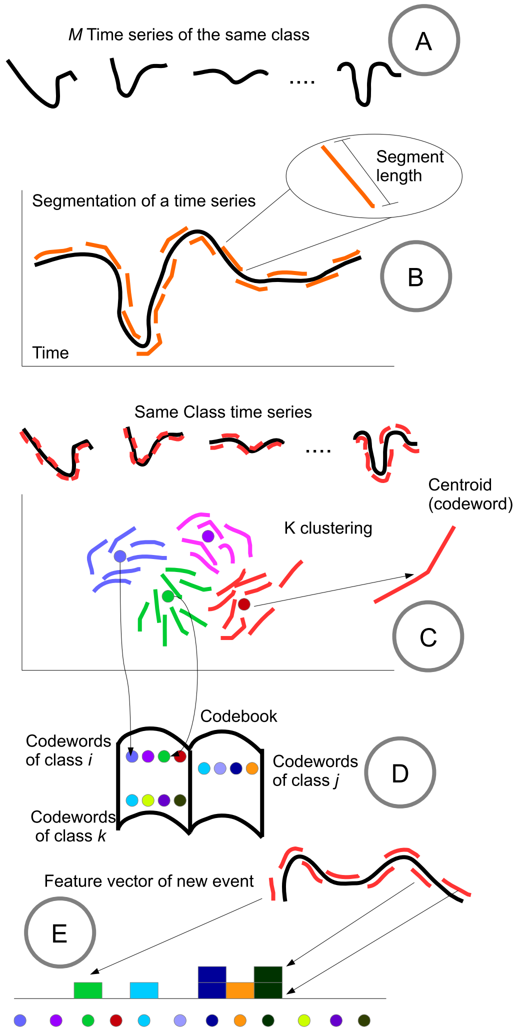


Fig. 5. Preprocessing procedure for the accelerometer data. A) The segmentation procedure is applied to all the training time series of a class. B) A time series is segmented, creating contiguous fragments of size n. C) With segments from all times series, $k-$means is applied to cluster similar signals, the resultant $k-$centers are considered as codewords. D) Since all classes went through the same procedure, a codebook is created with all the codeword sets. E) To create the feature vector, a time series is first segmented, then each segment is compared (using Euclidean distance) to every one of the codewords in the codebook, and it will be represented with the codeword that had shown the minimum distance, this way a histogram is created when all segments are compared, this histogram is the feature vector that is fed to the classifier.

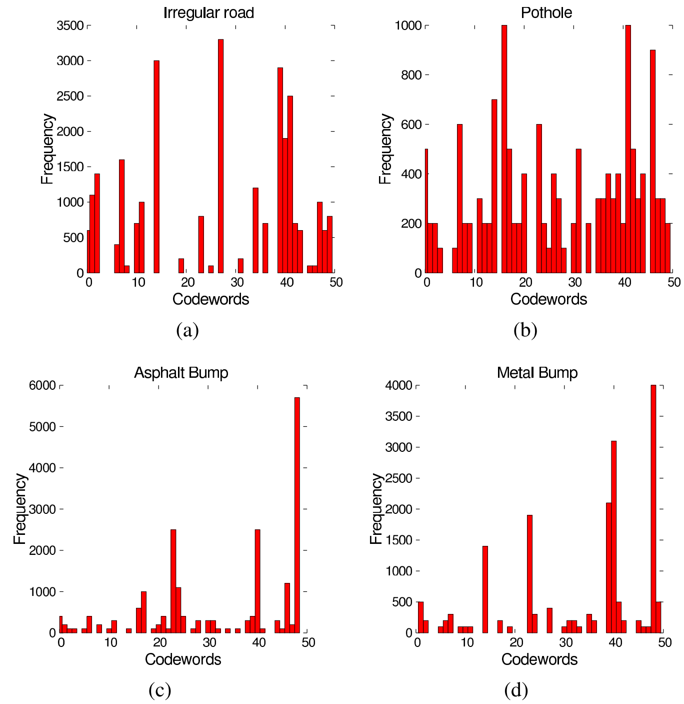


Fig. 6. BoW representation of five signals from different categories.

Learning to rank methodologies where the Bag of Words technique showed to better capture patterns within data. Additional discussion is presented in section 5.4. Figure 5 shows a graphical description the BoW formulation as applied to accelerometer data.

Under the BoW formulation, a codebook formed by prototypical patterns has to be learned. In the case of accelerometer data we propose learning a codebook as follows. Please note that in this work we focused in the preprocessing of the z-axis of the vehicle, since all anomalies directly impact this axis. Also note that acceleration in this axis can be obtained by reorienting the smartphone's readings, finding two angles and applying a rotation matrix [22]. Assume we have accelerometer measurements labeled with $C-$categories. For each of the $C-$categories we segment all of the training measurements (sensed data across time) into segments according to a predefined window size, n. The segmentation can be performed with overlap (i.e., using a sliding window over the temporal axis), or without it. Every obtained segment is represented by a descriptor. In our case, it comprises the n raw values of the signal, and the mean, variance, minimum and maximum of the raw values. In this way, every signal segment is associated to a vector representation. In a second stage, all of the descriptors corresponding to measurements from one class, are clustered with $k-$means into $k-$clusters. The centers of the $k-$clusters are taken as codewords. The process is repeated for all of the classes. As a result, we have a codebook formed by codewords extracted from samples of the different categories.

The assumption of BoW is that the content of objects can be effectively described by the distribution of prototypical patterns (codewords) appearing in the object. Therefore, accelerometer measurements are coded accordingly. Each represented segment is compared to the elements of the codebook, and segments are replaced by its closest codeword. Next, the signal is represented by an histogram that indicates the frequency of occurrence of codewords in it (see Figure 5E). This is the feature vector that classifiers use to recognize the class of the event.

Figure 6 shows the BoW representation of four signature events that belong to different categories (as a reference, the BoW representation for normal road is shown in Appendix A). It can be seen from this figure, that the BoW representation effectively captures discriminative patterns that simplify the task of the classification models: *different distribution of codewords are observed for the different categories*. Besides, the representation has a fixed length (all feature vectors have the same dimensionality), so that traditional (i.e., static) classifiers

TABLE III
APPROACHES EMPLOYED FOR CLASSIFICATION AND SPECIFIC CONFIGURATIONS

Method	Configuration parameters
ANN	layers = 2, hidden units = 25, function=softmax, Beta= 0.1, Momentum= 0.9
SVM	Kernel=Linear, C=1.0, shrinking=True, tol=0.001, random state=None
DT	max depth=None, min samples split=1,random state=0
RF	n estimators=10, max depth=None,min samples split=1, random state=0
NB	alpha=1.0, fit prior=True, class prior=None
KR	alpha=1.0, fit intercept=True, normalize=False, copy X=True, max iter=None, tol=0.001, class weight=None, solver='auto'
KN	n neighbors=5, weights='distance', algorithm='auto', leaf size=30, p=2, metric='minkowski', metric params=None

TABLE IV
MULTI CLASS DATA SETS CREATED TO COMPARE THE CLASSIFIERS

Data set	Description
BUMPS	2 classes: metal bumps and asphalt bump.
BUMPY ROAD	2 classes: comprises metal and asphalt bump into one class and regular road in another class.
BUMPY ROAD DETAILED	3 classes: metal bump in one class, asphalt bump into one class and regular road in another class.
ANOMALIES	4 classes: metal bump in one class, asphalt bump into one class, potholes into one class and regular road into another class.
ALL FOUR	4 classes: comprises metal and asphalt bump into one class, potholes in one class, regular road in one class and worn out road into another class.
ALL FIVE	5 classes: metal bump in one class, asphalt bump into one class, potholes in one class, regular road in one class and worn out road into another class.

can be applied, and, in our domain, the representation is not sparse and low dimensional. A limitation of the BoW formulation is that it disregards the temporal information explicitly, however, as shown in the next section, this limitation does not result in degraded performance when classifying RSDs. Finally, please note that the computational expensive procedures associated to BoW (codebook generation, and classifier training) are performed offline, and the time required to make predictions for a signal is negligible.

V. EXPERIMENTS AND RESULTS

In order to determine the usefulness of the BoW representation for classification of RSDs in the released data set, we applied and compared seven popular Machine Learning Techniques to tackle the problem. These techniques are representative of the wide range of algorithms in the field, comprising both linear and non linear models [23], [24]. The methods considered here are: Artificial Neural Networks (ANN), Support Vector Machines (SVM), Decision Trees (DT), Random Forrest (RF), Naive Bayes Classifier (NB), K-nearest neighbors (KNN) and Kernel Ridge (KR). For the first 6 methods we used the implementations from the *scikit-learn* Python-based package [25], whereas for KR we used the matlab® implementation from CLOP [26]. Table III shows the parameter configurations for each model. These parameters were the ones that resulted in better performance after preliminary experiments.

In our experiments, 60% of data were used for training and 40% for testing. We repeated this strategy 30 times per method and report average performance, this is to discard randomness effects. To assess the performance of the different methods we calculated accuracy, defined as (TP+TN)/(TP+TN+FP+FN),[3] and the Area Under the ROC Curve (AUC), which may be interpreted as the probability of classifying a random positive instance higher than a random negative instance. A perfect AUC score would be 1.0.

For our experiments, we considered different subsets of the data set described in Section III. Our motivation for this was two-fold. On the one hand we wanted to resemble most of the data sets already proposed in literature, this is to compare the performance of our methods with previous work. On the other hand, we wanted to measure the capabilities of the methods to tackle the sub classification problem (i.e., fine grained classification). Furthermore, the considered partitions provide evidence on the usefulness of the data set we propose. Table IV presents the data sets that were created and their description.

As could be seen in Section 4, the BoW representation has to be calibrated by adjusting its parameters. In Section V-B, we detail the parameter searching process. Appendix B presents the configuration of the parameters for the best classifier. In all cases, the codebook was constructed from the training data, which allow to generalize the presented results. The remainder of this section reports the experimental results.

A. Binary Classification

This section evaluates the performance of the proposed methodology when facing binary classification problems, that comprise fine-grained (e.g., BUMPS) and coarse-grained (BUMPY ROAD) problems, see Table IV. We distinguish fine from coarse grained (usual problem) because there are scenarios where a method capable of discriminating between similar events is desirable. This is the case for the events *Asphalt bump* and *Metal bump* which are caused by the presence of similar artifacts on the road, even some works in literature [5], [8], [11] report these two events under the same category *bumps*. We evaluated the performance of the 7 classifiers in this binary classification problem called BUMPS. The results, presented in Table 5, show a remarkable performance by almost all classifiers: all of them above 90% and the most effective methods are close to 100% of correct classifications. These results show the effectiveness of the BoW representation, suggesting that, with this formulation, sub classification of bumps is possible, which may open new research directions to fine-grained classification for all events.

In the BUMPY ROAD data set, the class *Bumps* now comprises classes *Asphalt Bump* and *Metal Bump*. Table 6 presents the results when the task is to discriminate BUMPS from regular road (coarse problem). The results are similar

[3]TP, TN, FP and FN, stand for True Positive, True Negative, False Positive and False Negative, respectively.

TABLE V
PERFORMANCE ON DATA SET BUMPS

Event	Asphalt Bumps	
	Accuracy %	AUC
ANN	99.3	0.993
SVM	98.3	0.983
DT	97.7	0.977
RF	98.3	0.982
NB	98.9	0.989
KR	98.1	0.981
KNN	90.1	0.902

TABLE VI
PERFORMANCE ON DATA SET BUMPY ROAD

Event	Bumps	
	Accuracy %	AUC
ANN	99.9	0.998
SVM	99.7	0.997
DT	98.5	0.98
RF	99.3	0.99
NB	89.4	0.89
KR	99.9	0.99
KNN	99.7	0.99

TABLE VII
PERFORMANCE ON THE DATA SET BUMPY ROAD DETAILED

Class	Asphalt Bump		Metal Bump		Regular Road	
	Acc %	AUC	Acc %	AUC	Acc%	AUC
ANN	99.0	0.998	96.9	0.968	98.9	0.986
SVM	97.6	0.976	96.3	0.964	98.5	0.979
DT	97.0	0.968	92.5	0.92	94.2	0.933
RF	97.6	0.976	95.1	0.938	97.3	0.972
NB	96.4	0.971	87.9	0.882	86.7	0.871
KR	96.3	0.961	97.7	0.98	98.5	0.979
KNN	92.4	0.913	92.8	0.933	97.7	0.967

to those from Table 5: all of the classifiers showing high performance, being the ANN the one that obtained the highest score (virtually 100%). Although a fair comparison is not possible, the BUMPY ROAD data set is similar in spirit with the one used in the Nericell project [5]. Nericell reports that for a speed $\geq$ 25 km/h, their classifier achieved 41% and 8% for False Negative and False Positive rate, respectively. The ANN achieves in this study a 0.01% for FNR and a 0.0% for a FPR. Again, this result highlights the benefits of using the BoW representation.

Table VII presents classification scores on the BUMPY ROAD DETAILED data set when faced as binary classification. We can observe that is slightly easier to identify *Asphalt bumps* than *Metal Bumps*. A previous analysis on bump sub classification, made by Astarita *et al.* [8], based on high energy events suggests this possibility.

Table VIII presents the results on the ANOMALIES data set. The works of Pertunnen *et al.* [12] and Seraj *et al.* [15] employed data sets that comprise two classes: *Anomaly* and *Regular road*. Anomalies in those works were formed either by speed bump, large potholes, bumps, cobblestone or cracks, manholes, patches, etc. depending on each work. Perttunnen et al. reported a FPR of 3% and a FNR of 18%, whereas Seraj et al. scored 11% and 12% for FPR and FNR, respectively. In this case, if we consider the classification for *Regular Road class* against the rest, the ANN obtains 0.5% for FPR and 0.6% for FNR. Therefore, although the results are not directly comparable,[4] our method would obtain much better performance than the reference.

The results for the data set ALL FOUR are presented in table IX. In the work of Astarita *et al.* [8] they performed a classification in a road test site that was characterized by bumps, rough pavement and potholes, being a similar scenario to ALL FOUR data set. They report to detect 90% of bump events and a FPR for potholes of about 35%. In this case, the ANN has a detection score of about 93% and it presents a 1.8% for FPR for potholes.

Table X presents the results when all classes are analyzed in the data set ALL FIVE. In general, all classes are discriminated with very competitive scores. Even the classifier that obtained the lowest accuracy is comparable with the results from literature. To add some context take into consideration the work of Fazeen *et al.* [10], they applied their approach to a five-class data set composed of *bump, pothole, rough, smooth* and *uneven* classes, obtaining classification accuracies of 81.5%, 72.2%, 75%, 91.5% and 89.4%, respectively. In average, they have an 85.7% of accuracy, while the ANN achieves 93.8% on average.

In order to have an overview of the comparisons, Table XI summarizes the results of the best score found for each data set in this work and the corresponding one from literature. Appendix C presents False Positive/Negative Rates for the best classifier.

B. Sensitivity Analysis

In the previous section we reported the performance obtained by the best configuration of the BoW representation. This section elaborates on the performance of the BoW representation for different parameter configurations. We considered different values for the following parameters: the length of the segment n (5, 8, 10, 12); the number of codewords k (10, 20, 30, 40, 50); using or not overlap (i.e. using sliding window or non overlapping segments); and the amount of training data used to building the codebook t (40%, 50%, 60%). There are 120 possibilities for combinations of these values, therefore the algorithm needs to be run to find which set of parameter better work for it. In this analysis, we ran the method 30 times for each combination, this with the purpose to average emergent behavior of elements present within this methodology, e.g., selection of different centroids in the k-means algorithm that may influence the generation of the codebook. In the remainder we report a summary of the results.

Figure 7a shows the behavior of three classifiers (ANN, SVM and NB) when applied to the BUMPS data set. For each classifier the accuracy of the two classes are shown as a function of the parameters combinations. Interestingly, there are sections where apparently an accuracy-decreasing pattern appears, these regions spans the indexes of parameter configurations 20-40, 60-80 and 100-120, which correspond to non overlapping segments. Figure 7b shows the same analysis over the BUMPY ROAD data set, this time showing the

[4] Please note that, as previously mentioned, data sets used in previous works were not available.

TABLE VIII

ACCURACY AND AUC OBTAINED BY EACH APPROACH FOR ALL CLASSES IN THE DATA SET ANOMALIES

Class	Asphalt Bump		Metal Bump		Potholes		Regular Road	
	Accuracy %	AUC	Accuracy %	AUC	Accuracy%	AUC	Accuracy%	AUC
ANN	95.9	0.949	93.6	0.912	93.8	0.92	99.5	0.99
SVM	93.5	0.916	91.7	0.893	93	0.913	99	0.984
DT	92.2	0.904	91.4	0.895	84	0.797	96.5	0.954
RF	93.5	0.907	94.6	0.918	88.7	0.832	98.5	0.979
NB	86.9	0.896	86.8	0.872	78	0.749	86	0.852
KR	93.5	0.921	92.1	0.894	94.9	0.935	99	0.983
KNN	90.1	0.839	88.8	0.846	90	0.88	98	0.962

TABLE IX

ACCURACY AND AUC OBTAINED BY EACH APPROACH FOR ALL CLASSES IN THE DATA SET ALL FOUR

Class	Bumps		Potholes		Regular Road		Worn out Road	
	Accuracy %	AUC	Accuracy %	AUC	Accuracy%	AUC	Accuracy%	AUC
ANN	93.1	0.906	95.4	0.937	92.3	0.896	87.0	0.824
SVM	89.5	0.870	95.2	0.935	90.8	0.878	81.7	0.743
DT	90.7	0.875	85.9	0.82	87.68	0.833	82.5	0.777
RF	93.7	0.913	91.0	0.838	90	0.856	85.3	0.809
NB	83.3	0.833	81.4	0.784	80.7	0.796	72.9	0.728
KR	90.8	0.862	94.0	0.934	92.2	0.89	82.1	0.728
KNN	89.7	0.87	90.4	0.893	91.7	0.861	87.5	0.822

TABLE X

ACCURACY AND AUC OBTAINED BY EACH APPROACH FOR ALL CLASSES IN THE DATA SET ALL FIVE

Class	Metal Bump		Asphalt Bump		Potholes		Regular Road		Worn out Road	
	Accuracy %	AUC	Accuracy %	AUC	Accuracy%	AUC	Accuracy%	AUC	Accuracy%	AUC
ANN	93.6	0.902	96.5	0.944	94.7	0.92	94.1	0.912	90.1	0.85
SVM	91.0	0.872	94.5	0.919	94.7	0.919	92.4	0.884	85.6	0.746
DT	90.5	0.864	92.4	0.886	83.1	0.747	91.0	0.855	84.6	0.761
RF	94.3	0.898	94.4	0.924	88.6	0.742	92.4	0.862	87.7	0.798
NB	84.1	0.837	88.5	0.91	76.1	0.725	85	0.822	82.9	0.801
KR	91.9	0.857	93.5	0.902	93.7	0.898	93.2	0.888	86.0	0.734
KNN	89.5	0.845	91.0	0.82	89.6	0.87	92.9	0.857	89.4	0.8

TABLE XI

SUMMARY OF COMPARISONS BETWEEN THE BEST CLASSIFIER REPORTED IN THIS WORK AND WORKS FROM LITERATURE

Work	**Metric/score**	**This work(ANN)**	**Data set**
Nericell [5]	FNR=41%,FPR=8%	FNR=**0.01%**,FPR=**0.0%**	BUMPY ROAD
Pertunnen et al. [12]	FNR=18%,FPR=3%	FNR=**0.6%**,FPR=**0.5%**	ANOMALIES
Seraj et al. [15]	FNR=12%,FPR=11%	FNR=**0.6%**,FPR=**0.5%**	ANOMALIES
Astarita et al. [8]	ACC=90%,FPR=35%	ACC=**93%**,FPR=**1.8%**	ALL FOUR
Fazeen et al. [10]	ACC=85.7%	ACC=**93.8**	ALL FIVE

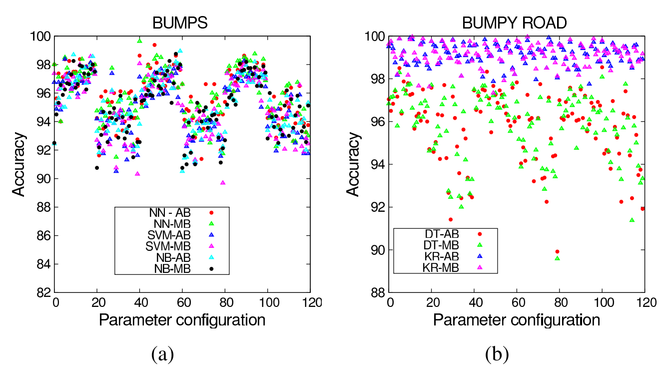


Fig. 7. Average accuracy over data set BUMPS (a) and data set BUMPY ROAD (b) as a function of parameter configuration for using different classifiers: NN (Artificial Neural Networks), SVM (Support Vector Machine), NB (Naive Bayes), DT (Decision Tree) and KR (Kernel Ridge).

DT and KR classifiers for *Asphalt* and *Metal bumps* classes. The pattern is more evident for DT, whereas the KR exhibits a more robust performance.

Figure 8 shows accuracy of all classifiers on *Potholes* and *Worn out* road in the data set ALL FIVE. The decrease in accuracy spans the same indexes as in previous figures, strongly suggesting that an overlapping strategy is the one to be used when performing this Bag of words pre-processing. Although this point needs to be further investigated, a possible reason that could explain the success of this overlapping strategy is that it captures time-dependencies between segments of the time-series, thus enriching the original BoW proposal which neglected this situation.

If we just focus on the results when the overlapping strategy is employed, the results suggest two points that we would like to remark. Apparently there is no influence in the amount of

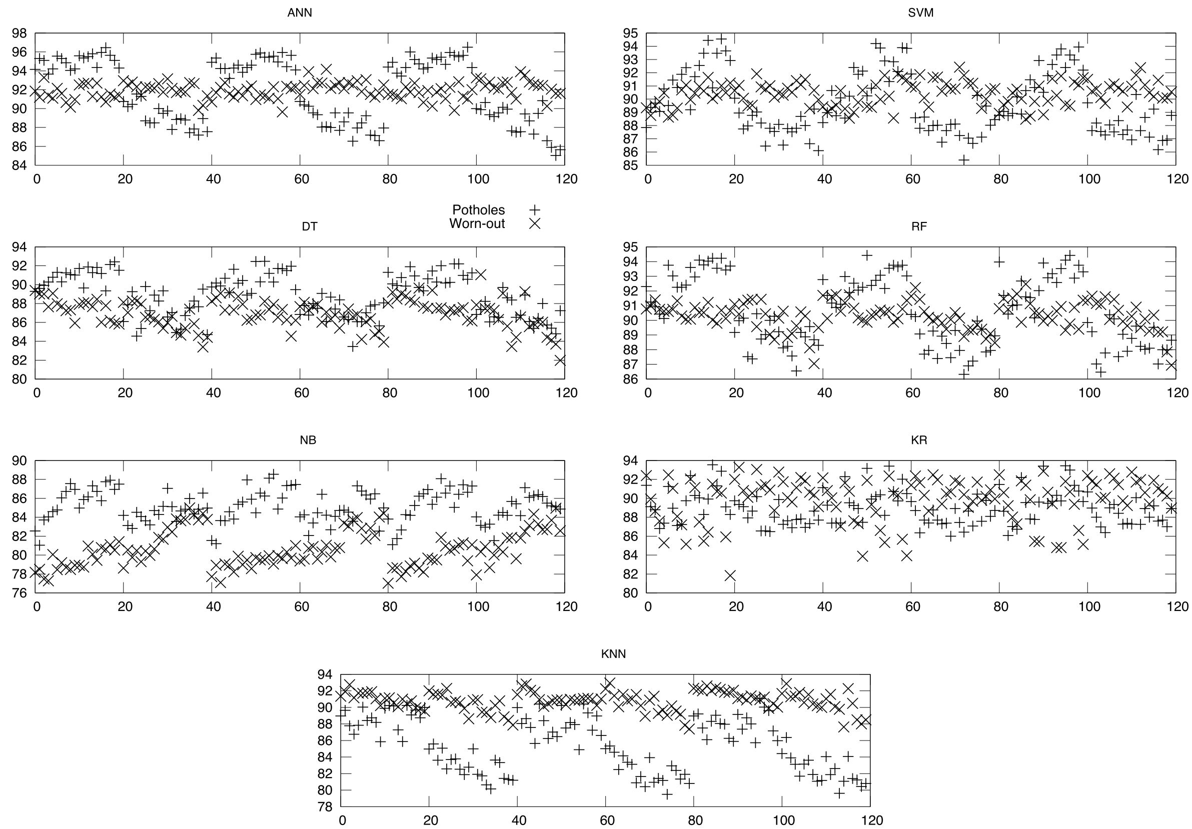


Fig. 8. Accuracy of all algorithms with respect to every possible configuration in the pre-processing stage for the ALL-FIVE data set, only considering classes *Potholes* and *Irregular road*.

TABLE XII
ACCURACY OBTAINED BY EACH APPROACH IN ALL DATA SETS

Class	Metal Bump (%)	Asphalt Bump (%)	Potholes (%)	Regular Road (%)	Worn out Road (%)
ANN	92.05	95.9	94.55	94.15	89.85
SVM	87.15	92.33	95.27	91.75	80.53
DT	90.45	92.6	80.85	93.47	84.33
RF	93.73	94.18	88.18	93.8	87.18
NB	76.05	82.48	66.78	83.33	82.0
KR	89.6	91.93	95.22	94.42	84.4
KNN	91.02	90.18	90.47	92.55	89.2

data used to create the codebooks, which is very interesting since we may be biased to think that a phenomenon such over/under fitting may be possible. The other point is that there is no significant influence of the two other parameters (codeword length and codebook size) in the accuracy scores, as stated by Wang *et al.* [20] in their original contribution.

C. Multi-class Classification

An interesting question is to compare the current results against those from multi-class classification schemes. Table XII presents these results, where a classifier now chooses among all the classes, the one with the highest probability. As we can observe (having in perspective table X), in general, binary classification outperforms a multi-class scheme.

TABLE XIII
PERFORMANCE OF A COUNTING STRATEGY APPLIED AT PREPROCESSING STAGE

BUMPY ROAD	BUMPY ROAD DETAILED	ANOMALIES	ALL FOUR	ALL FIVE
83%	65%	46.5%	53.75%	45%

D. Comparison With Alternative Techniques

According to the way we preprocess the time series, we can think of an straightforward classification procedure, bypassing the Machine Learning approaches: for a given time series, we can select the class that had the highest number of codewords employed. This is a fair experiment that would help unveil the role that machine learning techniques play in this problem since no pattern finding strategy is being used. Table XIII presents the percentage of correct classification for all data sets when this simple idea is followed. These results highlight the advantages of complementing the BoW approach with a more sophisticated pattern extraction algorithm.

Another obvious option would be to apply the Dynamic Time Warping (DTW) algorithm [27] to compare time series. To reduce the computational cost of applying DTW, we used the approximation coefficients of a Discrete Wavelet Transform (DWT) [28]. Following this idea, a third level DWT was calculated using PyWavelets [29] with a Daubechies D4/db2 wavelet. The DWT approximation coefficients were

TABLE XIV
AVERAGE DIFFERENCE BETWEEN PAIRS OF SAMPLES OF ALL FIVE CATEGORIES (ARBITRARY UNITS)

Class	**Asphalt Bump**	**Metal Bump**	**Pothole**	**Regular Road**	**Worn out Road**
Asphalt Bump	0.3186	0.3640	0.3885	0.4285	0.4166
Metal Bump	0.3640	0.3821	0.4078	0.4436	0.4309
Pothole	0.3885	0.4078	0.4083	0.4735	0.4533
Regular Road	0.4285	0.4436	0.4735	0.4116	0.4265
Worn out Road	0.4166	0.4309	0.4533	0.4265	0.4293

normalized before calculating the difference between pairs. In total, the difference between more than 250,000 pairs of signals were calculated. Table XIV shows the average difference found with DTW between pair of samples of the five categories previously discussed. The message conveyed by this result is very striking since no one of the classes better resembles itself, thus discarding the idea of just using this algorithm for classification purposes.

E. Discussion

Previous works in literature based on the assumption that current online processing is still limited have focused their efforts in proposing threshold heuristics. Nonetheless, these strategies have shown their own limitations given the high rate of FP and FN reported. This fact has motivated the emerging appearance of Machine Learning techniques, but until now, no study had employed a wide range of these approaches to measure their real capabilities to handle this problem. For the binary classification, it is remarkable that most of the classifiers showed very promising results, clearly outperforming the best results reported. This last point suggests that the preprocessing stage, Bag of Words, plays a key role to capture patterns for all the classes.

Regarding the classification models, ANN was the approach with the highest accuracy and AUC scores. Although there are very competitive algorithms like the SVM and the KR that obtained very good performance. When we performed a sensitivity analysis for the Bag of Words technique, it was observed that most of the best classification results were obtained when overlapping segments were employed. We can hypothesize that overlapping may contribute to alleviate noise, presented in the signals given their heterogeneity. This could be a possibility, since time-dependencies are captured when using overlapping segments. It is interesting that very high classification scores were obtained and this may extend previous application perspectives of Bag of Words technique, since Wang *et al.* [20] proponents of this preprocessing stage textually claimed in their Discussion section: "... *the Bag-of-words representation may be ineffective to represent short time series, which is mainly due to the limitation that the bag-of-words representation cannot extract enough meaningful and discriminative local segments from short sequences*". This

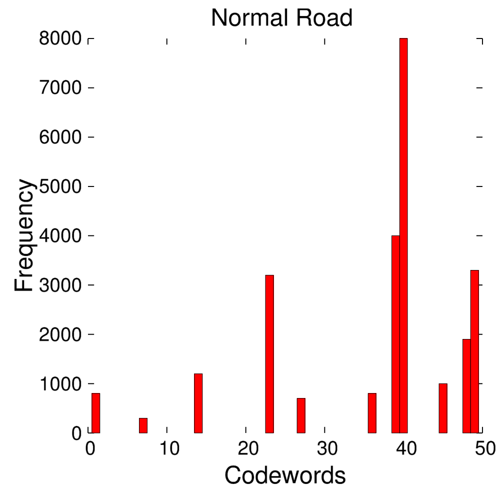


Fig. 9. BoW representation of the class Normal road.

TABLE XV
BEST PARAMETERS OF BoW FOR THE BUMPS DATA SET

Bumps	Asphalt Bump			
Algorithm	%Data	Segment Length	Overlpping	CodeWords per Class
ANN	50	5	1	10

TABLE XVI
BEST PARAMETERS OF BoW FOR THE BUMPY DATA SET

Bumpy Road	Bumps			
Classifier	%Data	Segment Length	Overlapping	CodeWords per Class
ANN	40	5	0	20

same analysis shed some light on the role that the amount of data to construct the codebook plays, since it was found no important difference when comparing different values for this parameter.

To promote further analysis and comparisons with other techniques, we make publicly available all data sets and pseudocode of the *Bag of words* at http://www.accelerometer.xyz/datasets.

VI. CONCLUSIONS

This work tackles the problem of classifying Roadway Surface Disruption from a Machine Learning standpoint. For the experimental section, several datasets similar to those previously used in the literature were created. These datasets are composed of accelerometer signals collected in several places within the cars, using different cars and drivers. Altogether, making it the largest and most heterogenous data set reported. To create the feature vectors, we represented accelerometer signals with a novel adaptation of the *Bag of Words* (BoW) approach. In order to have a better grasp of this approach, a sensitivity analysis was performed, resulting in a performance boost with the usage of a new overlapping strategy. Further analysis is needed to completely understand the reasons for the success of this overlapping strategy. Another open question that needs to be addressed is related to the computational requirements of BoW. So far, we have considered that BoW could reside on a server in the cloud, avoiding intensive

TABLE XVII
BEST PARAMETERS OF BoW FOR THE ANOMALIES DATA SET

	Asphalt Bump				Metal Bump				Potholes				Regular Road			
Classifier	%Dat	SLe	Ov	Wor	%Dat	SLe	Ov	Wor	%Dat	SLe	Ov	Wor	%Dat	SLe	Ov	Wor
ANN	60	10	1	10	40	12	1	20	40	12	1	40	60	10	1	10

TABLE XVIII
BEST PARAMETERS OF BoW FOR THE BUMPY ROAD DETAILED DATA SET

	Asphalt Bump				Metal Bump				Regular Road			
Classifier	%Dat	SLe	Ov	Wor	%Dat	SLe	Ov	Wor	%Dat	SLe	Ov	Wor
ANN	40	12	1	40	60	8	1	30	60	8	1	30

TABLE XIX
BEST PARAMETERS OF BoW FOR THE ALL FOUR DATA SET

	Bumps				Potholes				Regular Road				Worn out road			
Classifier	%Dat	SLe	Ov	Wor	%Dat	SLe	Ov	Wor	%Dat	SLe	Ov	Wor	%Dat	SLe	Ov	Wor
ANN	60	10	1	10	60	12	1	10	60	5	1	30	60	5	0	20

TABLE XX
BEST PARAMETERS OF BoW FOR THE ALL FIVE DATA SET

	Metal Bump				Asphalt Bump				Potholes				Regular Road				Worn out Road			
Class	%Dat	SLe	Ov	Wor	%Dat	SLe	Ov	Wor	%Dat	SLe	Ov	Wor	%Dat	SLe	Ov	Wor	%Dat	SLe	Ov	Wor
ANN	60	12	1	10	60	12	1	40	50	10	1	10	50	8	0	20	60	5	1	30

processing on the smartphone, but experimental analysis needs to be performed to evaluate the alternative of online processing in the users' devices.

Although a direct and fair comparison with algorithms from the literature is not possible, seven classifiers were run onto the new datasets. They showed very competitive performance, both in accuracy and FP/FN rates, in relation to state of the art algorithms. The classifier that obtained the best results was the Artificial Neural Network, which positions itself as the classifier with the highest accuracy score reported for this problem.

As a further work we would like to: (1) create a larger data set that could pose bigger challenge for these classifiers, (2) fully characterize the BoW approach and its new overlapping strategy, (3) have an implementation of the algorithms from literature in order to run over the same data set, thus allowing a fair comparison.

APPENDIX A
HISTOGRAM

For comparison purposes, Figure 9 shows the BoW representation for the class of Normal (Regular) road.

APPENDIX B
PARAMETERS OF THE BAG OF WORDS PREPROCESSING STAGE

Tables XV, XVI, XVII, XVIII, XIX and XX present the particular set of parameters that were used in the BoW preprocessing stage for the Artificial Neural Network in the experimental section. What is presented is: the amount of data used to build the codebook (%Dat), length of segment (SLe), if overlapping was performed or not (Ov), and the number of codewords per class or number of centroids (Wor).

TABLE XXI
FALSE POSITIVE AND FALSE NEGATIVE RATES (%) FOR BUMPS DATA SET

Event	Asphalt Bump	
	FPR	FNR
ANN	0.075	0.625

TABLE XXII
FALSE POSITIVE AND FALSE NEGATIVE RATES (%) FOR BUMPY ROAD DATA SET

Event	Bumps	
	FPR	FNR
ANN	0	0.01

TABLE XXIII
FALSE POSITIVE AND FALSE NEGATIVE RATES (%) FOR BUMPY ROAD DETAILED DATA SET

Class	Asphalt Bump		Metal Bump		Regular Road	
	FPR	FNR	FPR	FNR	FPR	FNR
ANN	0.54166667	0.45833333	1.8	1.3	1.1	0

APPENDIX C
FALSE POSITIVE/NEGATIVE RATES

Tables XXI, XXII, XXIII, XXIV, XXV and XXVI, complement the performance metrics of the Artificial Neural Network by presenting the percentage of False Positive and False Negative Rates, FPR and FNR, respectively, for every dataset.

TABLE XXIV
FALSE POSITIVE AND FALSE NEGATIVE RATES (%) FOR ANOMALIES DATA SET

Class	Asphalt Bump		Metal Bump		Potholes		Regular Road	
	FPR	FNR	FPR	FNR	FPR	FNR	FPR	FNR
ANN	2.14375	1.95625	3.1375	3.2625	3.63125	2.56875	0.5	0.6

TABLE XXV
FALSE POSITIVE AND FALSE NEGATIVE RATES (%) FOR ALL FOUR DATA SET

Class	Bumps		Potholes		Regular Road		Regular Road	
	FPR	FNR	FPR	FNR	FPR	FNR	FPR	FNR
ANN	3.66875	3.23125	1.8	2.45625	4.69375	3.00625	6.875	6.125

TABLE XXVI
FALSE POSITIVE AND FALSE NEGATIVE RATES (%) FOR ALL FIVE DATA SET

Class	Metal Bump		Asphalt Bump		Potholes		Regular Road		Worn out road	
	FPR	FNR	FPR	FNR	FPR	FNR	FPR	FNR	FPR	FNR
ANN	3.225	3.175	1.625	1.875	2.425	2.875	3.7	2.2	5.55	4.35

ACKNOWLEDGEMENTS

The authors are very grateful to the anonymous reviewers who helped them to improve the contribution of this work.

REFERENCES

[1] I. Kertész, T. Lovas, and A. Barsi, "Photogrammetric pavement detection system," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 37, pp. 897–902, 2008.

[2] C. M. Jengo, D. Hughes, J. D. LaVeigne, and I. Curtis., "Pothole detection and road condition assessment using hyperspectral imagery," in *Proc. Annu. Conf. Amer. Soc. Photogramm. Remote Sens. (ASPRS)*, 2005, pp. 7–11.

[3] Y. Du, C. Liu, D. Wu, and S. Jiang, "Measurement of international roughness index by using Z -axis accelerometers and GPS," *Math. Problems Eng.*, vol. 2014, Mar. 2014, Art. no. 92898.

[4] J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden, and H. Balakrishnan, "The pothole patrol: Using a mobile sensor network for road surface monitoring," in *Proc. 6th Int. Conf. Mobile Syst., Appl., Services*, 2008, pp. 29–39.

[5] P. Mohan, V. N. Padmanabhan, and R. Ramjee, "Nericell: Rich monitoring of road and traffic conditions using mobile smartphones," in *Proc. 6th ACM Conf. Embedded Netw. Sensor Syst.*, 2008, pp. 323–336.

[6] G. D. De Silva, R. S. Perera, N. M. Laxman, K. M. Thilakarathna, C. I. Keppitiyagama, and K. De Zoysa, "Automated pothole detection system," in *Proc. Int. Conf. Adv. ICT Emerg. Regions*, Colombo, Sri Lanka, 2013.

[7] A. Mednis, G. Strazdins, R. Zviedris, G. Kanonirs, and L. Selavo, "Real time pothole detection using Android smartphones with accelerometers," in *Proc. Int. Conf. Distrib. Comput. Sensor Syst. Workshops (DCOSS)*, Jun. 2011, pp. 1–6.

[8] V. Astarita *et al.*, "A mobile application for road surface quality control: Uniqualroad," *Proc.-Soc. Behavioral Sci.*, vol. 54, pp. 1135–1144, Oct. 2012.

[9] V. Douangphachanh and H. Oneyama, "A study on the use of smartphones for road roughness condition estimation," *J. Eastern Asia Soc. Transp. Studies*, vol. 10, pp. 1551–1564, 2013.

[10] M. Fazeen, B. Gozick, R. Dantu, M. Bhukhiya, and M. C. González, "Safe driving using mobile phones," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 3, pp. 1462–1468, Sep. 2012.

[11] A. Sinharay, S. Bilal, A. Pal, and A. Sinha, "Low computational approach for road condition monitoring using smartphones," in *Proc. 1st Int. Conf. Intell. Infrastruct. 47th Annu. Nat. Conv. Comput. Soc. India CSI*, 2013.

[12] M. Perttunen *et al.*, "Distributed road surface condition monitoring using mobile phones," in *Proc. Int. Conf. Ubiquitous Intell. Comput.*, 2011, pp. 64–78.

[13] L. C. González, F. Martínez, and M. R. Carlos, "The citizen road watcher–identifying roadway surface disruptions based on accelerometer patterns," in *Proc. Int. Conf. Ubiquitous Comput. Ambient Intell.*, 2013, pp. 374–377.

[14] F. Martínez, L. C. González, and M. R. Carlos, "Identifying roadway surface disruptions based on accelerometer patterns," *IEEE Latin Amer. Trans.*, vol. 12, no. 3, pp. 455–461, May 2014.

[15] F. Seraj, B. J. van der Zwaag, A. Dilo, T. Luarasi, and P. Havinga, "RoADS: A road pavement monitoring system for anomaly detection using smart phones," in *Proc. Int. Workshop Modeling Social Media*, 2014, pp. 128–146.

[16] H. R. Eftekhari and M. Ghatee, "An inference engine for smartphones to preprocess data and detect stationary and transportation modes," *Transp. Res. C, Emer. Technol.*, vol. 69, pp. 313–327, Aug. 2016.

[17] D. Figo, P. C. Diniz, D. R. Ferreira, and J. M. P. Cardoso, "Preprocessing techniques for context recognition from accelerometer data," *Pers. Ubiquitous Comput.*, vol. 14, no. 7, pp. 645–662, 2010.

[18] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, Oct. 2003, pp. 1470–1477.

[19] S. Manchala, V. K. Prasad, and V. Janaki, "GMM based language identification system using robust features," *Int. J. Speech Technol.*, vol. 17, no. 2, pp. 99–105, 2014.

[20] J. Wang, P. Liu, M. F. H. She, S. Nahavandi, and A. Kouzani, "Bag-of-words representation for biomedical time series classification," *Biomed. Signal Process. Control*, vol. 8, no. 6, pp. 634–644, 2013.

[21] J. Lin, R. Khade, and Y. Li, "Rotation-invariant similarity in time series using bag-of-patterns representation," *J. Intell. Inf. Syst.*, vol. 39, no. 2, pp. 287–315, 2012.

[22] M. R. Carlos, L. C. González, F. Martínez, and R. Cornejo, *Evaluating Reorientation Strategies for Accelerometer Data From Smartphones for ITS Applications*. Cham, Switzerland: Springer, 2016, pp. 407–418.

[23] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. Hoboken, NJ, USA: Wiley, 2012.

[24] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*. New York, NY, USA: Springer, Feb. 2013.

[25] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.

[26] A. Saffari and I. Guyon, "Quick start guide for CLOP," Graz Univ. Technol. Clopinet, Tech. Rep., May 2006.

[27] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *Proc. KDD Workshop*, vol. 10. Seattle, WA, USA, 1994, pp. 359–370.

[28] A. Graps, "An introduction to wavelets," *IEEE Comput. Sci. Eng.*, vol. 2, no. 2, pp. 50–61, May 1995.

[29] F. Wasilewski. (2010). *Pywavelets: Discrete Wavelet Transform in Python*. [Online]. Available: http://www.pybytes.com/pywavelets

Luis C. González (M'–) received the Ph.D. degree from University of North Carolina at Charlotte in 2011. In 2011, he joined Universidad Autónoma de Chihuahua, in the north of Mexico, where he was a Founder and currently a Professor of the Graduate Program of Computer Engineering. He is a member of the National System of Researchers, National Council of Science and Technology of México. His current research interests are pattern recognition and combinatorial optimization problems.

Ricardo Moreno received the B.Eng. in software engineering from Universidad Autónoma de Chihuahua in 2014, where he is currently working toward the M.Eng. degree in computer engineering. His research interest includes Machine Learning Algorithms and Pattern Recognition.

Hugo Jair Escalante received the Ph.D. degree from Instituto Nacional de Astrofísica, Óptica y Electrónica in 2010. He has been a Titular Research Scientist with Instituto Nacional de Astrofísica, Óptica y Electrónica, since 2012. He has been the Director of ChaLearn, the challenges in machine learning organization since 2011, and also has been a member of the National System of Researchers, since 2010. His main interests are on machine learning and its application on computer vision and natural language processing.

Fernando Martínez received the master's degree in industrial electronic engineering from the Instituto Tecnologico de Chihuahua–Chihuahua, Mexico, in 1997 and the Ph.D. degree in computer science from University of Nottingham, Nottingham, U.K., in 2009. His research interests are mobile computing, sensor networks, and human–computer interactions.

Manuel Ricardo Carlos received the B.Eng. in computer systems engineering from Chihuahua Institute of Technology II in 2010 and the M.Eng. in software engineering from Universidad Autónoma de Chihuahua in 2015, where he is currently working toward the Ph.D. degree in engineering. His research interests include machine learning and mobile computing.