

# Identifying Roadway Surface Disruptions Based on Accelerometer Patterns

L. C. González, F. Martínez and M. R. Carlos

**Abstract**— In developed countries large-scale technology is used for the monitoring and management of road infrastructure. Sensor networks together with surveillance cameras help identify elements disturbing cars' mobility. On the contrary, in developing countries, the lack of this level of infrastructure makes the task of roads' maintenance and traffic control challenging. The presence of roadway surface disruptions (RSDs), in particular, affects the economy and reputation of individuals and companies. Urban computing strategies can be adopted to collect data about the presence of RSDs and be applied to monitor traffic related issues. The implementation of such level of navigation assistance could help save freight and money to companies and authorities, and even reduce cars' accidents. In this direction, current mobile technologies can help with the identification and location of road imperfections alerting drivers about alternative routes. In this paper we present our work that seeks to enable citizens' cars as road watchers. By means of the mobile phones' acceleration sensing capabilities we are identifying and tagging the presence of RSDs. Using Android-based devices situated on the copilot floor side of a car, 5 Mbytes of road information has been collected. We run a series of experiments aiming to differentiate acceleration patterns associated to potholes, speed bumps, metal humps and rough roads. Currently, the classification of disruptions is being experimented with techniques from the field of Machine Learning (ML) such as artificial neuronal networks and logistic regression. Classification of individual events is over 86% of accuracy that is competitive with those reported in the literature. In this work we provide the first public dataset that could be used by other researchers to offer more insight in this problem.

**Keywords**— road surface disruptions, RSD, mobile computing, mobile sensing, machine learning algorithms.

## I. INTRODUCCIÓN

**B**ACHES, falta de señalización en topes y boyas, cruce de vías de tren en malas condiciones, grietas en pavimentos y tapas de registros de desagües pluviales removidas, conocidas de aquí en adelante como RSDs, "Roadway Surface Disruptions", son algunos elementos que alteran el flujo vehicular de los sistemas de comunicación vial, generando caos en el tráfico y en el menor de los casos la frustración de los conductores.

En los países desarrollados existe un gran despliegue de tecnologías de vigilancia y monitoreo enfocadas a la supervisión y administración de las vías de comunicación terrestres. A lo largo de estas se cuenta con cámaras de

vigilancia con las cuáles es posible observar elementos que están alterando la movilidad de los automotores. Así mismo existen nodos de sensores a través de los cuales es posible identificar e incluso pronosticar, el impacto de las condiciones ambientales sobre caminos y carreteras. En los países en desarrollo, sin embargo, la administración de la infraestructura carretera es todo un reto. En algunos países el nivel de tecnología utilizada para estos propósitos es mínimo, lo cual reduce las posibilidades de contar no solo con adecuados sistemas de supervisión sino además con una atención ágil de los problemas que están afectando el tránsito de vehículos.

La aparición de RSDs, en los países en desarrollo, puede llegar a niveles caóticos. Los trabajos de reparación de baches puede demorar semanas e incluso meses. Los trabajos de repavimentación pueden estarse realizando en horas pico de tráfico. El mantenimiento de señalizaciones, e.g. reductores de velocidad, pueden incluso tardar meses. Topes y boyas pueden aparecer de forma repentina y sin autorización/regulación oficial si residentes de una zona habitacional así lo acuerdan. Zonas de pavimento irregular pueden existir si proveedores de servicios no aplican los parches adecuados al pavimento. Sin embargo, sin importar cual sea la naturaleza de la existencia de RSDs, estos afectan tanto a particulares como al gobierno y a la industria por igual. Por ejemplo, los daños causados a los automóviles impactan directamente la economía de la persona. La industria podría verse en la necesidad de alterar sus rutas de entrega de mercancía lo cual impacta los tiempos comprometidos con los clientes y por lo tanto su reputación. Para las agencias de gobierno la identificación y reparación de RSDs en las carreteras bajo su jurisdicción puede ser complicada y costosa.

Con un nivel de integración de tecnología adecuada podría ser posible cristalizar la visión futurista de los autos "inteligentes". Estos autos se presume formarán parte de la red de nodos de recolección de datos que darán soporte al sistema de toma de decisiones de los departamentos de administración de carreteras. Información acerca de condiciones ambientales, estado que guardan las carreteras, accidentes, delincuencia y tráfico es recolectada a través de estos nodos móviles, y utilizada para generar servicios que informan a los automovilistas de estos y otros posibles problemas en su trayecto, indicándole opciones de rutas alternativas para alcanzar su destino final en el menor tiempo posible [1].

El acceso a este tipo de autos por parte del ciudadano común podría tardar varias décadas. Afortunadamente, el nivel de desarrollo tecnológico en el mercado de las tecnologías emergentes nos brindan la oportunidad de experimentar con algunos escenarios funcionales de los autos del futuro. Los teléfonos inteligentes disponibles en el mercado tienen

L. C. González, Universidad Autónoma de Chihuahua, Chihuahua, México, gonzalezgurrola@gmail.com

F. Martínez, Universidad Autónoma de Chihuahua, Chihuahua, México, fmartine@uach.mx

M. R. Carlos, Universidad Autónoma de Chihuahua, Chihuahua, México, p168786@uach.mx

integrados sensores de luz y temperatura ambiental, humedad, presión atmosférica, acelerómetro y giróscopio, tecnología con la cual es factible desarrollar aplicaciones en el contexto de “Urban Computing” [2]. La literatura nos muestra algunos ejemplos en los cuáles el ciudadano se muestra como un nodo recolector de información y su participación sea esta implícita o explícita puede apoyar en el desarrollo urbano de la localidad. Por ejemplo, los gobiernos pueden contar con información de sitios y zonas de interés y utilizar esta información para proyectar el crecimiento comercial o habitacional de esas áreas de la ciudad. Contenido adicional reportado por el teléfono celular del ciudadano puede usarse para identificar niveles de contaminación en esas zonas de la ciudad [3], o conocer los problemas de tráfico dado los tiempos de traslado hacia o desde el trabajo/hogar [4]. El nivel de participación ciudadana puede extenderse hacia escenarios de identificación del estado de las carreteras y el correspondiente tráfico vehicular dada la presencia de RSDs.

Hasta este punto, podríamos resumir entonces que la presencia de RSDs afecta la administración de las vías de comunicación terrestres, en virtud de que la movilidad y el tráfico vehicular en muchos de los casos se ven alterados. Por otro lado, se pueden visualizar oportunidades de aplicación de la tecnología en el desarrollo de nodos móviles “inteligentes” que recolectan y procesan información del entorno en el que se encuentran. Esta información puede apoyar las estrategias de administración y mantenimiento de la infraestructura vial.

La organización del trabajo es la siguiente: en la sección 2 se presenta la revisión de algunos de los esfuerzos que la comunidad científica ha realizado para la implementación de nodos móviles enfocados a la identificación de RSDs, resaltando sus limitaciones. En la sección 3, se presentan los recursos técnicos y tecnológicos que dan soporte al presente proyecto. En la sección 4, se describe la contribución de este trabajo en el área de la identificación “inteligente” de RSDs; dispositivos móviles recolectan datos de RSDs y se aplican técnicas de aprendizaje de máquina para clasificar baches, topes, boyas y pavimento irregular. En la sección 5, se realiza una discusión de los resultados obtenidos de los experimentos realizados hasta el momento y finalmente, en la sección 6 se ofrecen las conclusiones y el trabajo futuro de este proyecto.

## II. TRABAJO PREVIO

El problema de la identificación automática y localización de baches, topes y boyas, entre otras irregularidades en los caminos y carreteras, se ha mantenido como un problema complejo y lleno de retos a pesar de haber sido abordado con una gran variedad de métodos. En esta sección desarrollamos una revisión de los trabajos que sobresalen en la identificación de RSDs, los cuales a nuestra opinión se pueden clasificar en dos grandes categorías tomando en consideración la gama de tecnología utilizada en sus experimentos. En la primera clase tenemos a trabajos basados en métodos ópticos y acústicos, mientras que en la segunda categoría a aquellos que hacen uso del acelerómetro integrado en algún tipo de sistema embebido, incluyendo los teléfonos inteligentes.

En la primera categoría, Jengo et al. [5] utilizan imágenes multi-espectrales, fotografía aérea de alta resolución y fotografía a nivel de piso sobre diferentes secciones de carreteras con el objetivo de identificar diferencias entre sus espectros de frecuencia, y a través de estas encontrar las irregularidades en el pavimento. Herramientas de procesamiento de imágenes basadas en árboles de regresión y clasificación (CART), y dedicadas al procesamiento de bandas espectrales (BANDMAX) son utilizadas para identificar diferencias a nivel pixel entre un continuo de imágenes de la carretera. Con estos métodos de análisis los autores obtuvieron hasta un 68% de precisión en los resultados. Kertész et al. [6] utilizan cámaras de video y proyectores de láseres para desarrollar un barrido de imágenes continuas de un carril completo de la carretera. El láser sirve de referencia para la imagen tomada del pavimento. Con técnicas de procesamiento de imágenes se realiza una comparación entre las imágenes de las cámaras. Las diferencias existentes tanto en dos dimensiones como en profundidad (o alteración de la línea de referencia láser) permiten detectar baches y objetos en el pavimento. En Mednis et al. [7] se desarrolló una plataforma vehicular para la grabación y análisis de audio proveniente de los baches a través de los cuales se conduce un automóvil. El audio se discretiza a 200 Hz. y a través del uso de umbrales basados en la amplitud de la señal se diferencia la existencia del bache, según se reporta, hasta con un 80 % de efectividad. Con excepción del último trabajo, las dos primeras propuestas hacen uso de tecnología especializada y su aplicación podría verse limitada debido a sus requerimientos de cómputo. El tercer trabajo requiere una sintonización *ad hoc* de los filtros digitales que ayudan en la clasificación de baches, lo cual podría limitar su aplicación en actividades de detección de otros tipos de RSDs.

En la segunda categoría se encuentran trabajos desarrollados por De Silva et al. [8] y Eriksson et al. [9], en los cuales se identifica el uso de sistemas embebidos que integran tecnología de acelerómetros. En el primero se cuenta con un acelerómetro de dos ejes (x,y) para detectar baches. La señal proveniente del acelerómetro se procesa primeramente con un filtro pasa-bajas y posteriormente se aplica un criterio de umbral para identificar los baches con hasta un 80% de éxito. En el segundo, se cuenta con acelerómetros de tres ejes (x,y,z) para diferenciar baches de otro tipo de irregularidades como el cruce de vías de ferrocarril o reparaciones realizadas a la carretera. Además de utilizar la aceleración en el eje z, se considera la relación de cambio entre este eje y el eje x, así como la relación entre la temporalidad del evento y la velocidad del automóvil. En UNiquALroad, Astarita et al. [10], reportan la clasificación de topes de concreto y topes de hule, utilizando dispositivos móviles que cuentan con acelerómetro. La señal es inicialmente pre-procesada utilizando filtros pasa-bajas y posteriormente, evaluada respecto a los cambios de energía sobre el eje z. El cuidadoso trabajo desarrollado para caracterizar el desempeño del acelerómetro respecto a las condiciones físicas de cada tipo de topes le permitió a este trabajo obtener un 95% de éxito en la detección de estas irregularidades. En el proyecto Nericell [11] se aborda la detección de irregularidades en las carreteras utilizando la resultante de los tres vectores de aceleración (x,y,z). Combinando esta información con la temporalidad de

del evento se establecen criterios para diferenciar topes de baches. En Fazeen et al. [12] se observa una extensión a las experiencias descritas anteriormente, ya que se propone un sistema que identifica no solo RSDs sino además el comportamiento del conductor. Para la detección de topes y baches se combina la información de los cambios en aceleración de los ejes z y x, así como el periodo de duración de estos eventos. La clasificación efectiva de eventos tiene un rango de entre 71 y 85%.

Como puede observarse, a pesar de ser un problema de clasificación bien identificado y planteado llama la atención el poco interés de la comunidad por utilizar algoritmos específicos de aprendizaje de máquina con posibilidades bien probadas para este tipo de problemas. En este trabajo se propone el uso de dos algoritmos para problemas de clasificación, una red neuronal y regresión logística. Otra aportación de este trabajo es la propuesta del primer dataset público que permitirá a la comunidad experimentar y contrastar procedimientos y técnicas para la identificación de RSDs.

### III. MATERIALES Y MÉTODOS

La fase de recolección de datos se realizó durante el verano de 2013 en 2 carros diferentes, un Nissan Sentra 2013 y un Nissan Tsuru 2001. En cada auto se realizaron diversos recorridos cada uno considerando el tránsito sobre avenidas, calles, bulevares y carreteras de la ciudad de Chihuahua, Chihuahua, México. La Fig. 1 muestra uno de estos recorridos cuya longitud es de 40 Kilómetros.

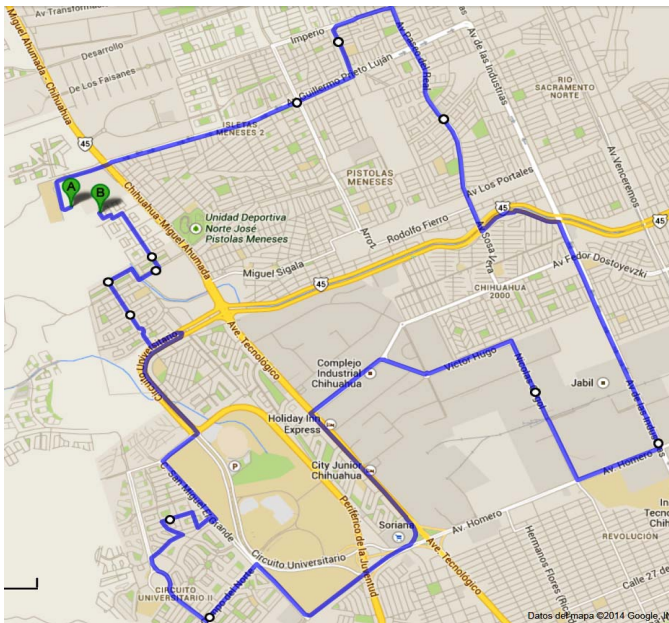


Figura 1. Una de las trayectorias de 40 Km. sobre la cual se recolectaron datos de RSDs.

Los dispositivos electrónicos que se utilizaron para la recolección de los datos fueron 1 teléfono celular HTC con sistema operativo Android 4.03, colocado en el piso del automóvil del lado del acompañante, en el cual se registraban datos de aceleración. Una tableta digital Motorola XOOM con sistema operativo Android 3.3 la cual primeramente funciona

como el control remoto del teléfono celular y permite sincronizar el registro de eventos, y en segundo lugar, registra las coordenadas GPS de la ubicación del evento.

#### A. Técnicas de aprendizaje computacional

##### 1) Red Neuronal

Las redes neuronales [13] son uno de los modelos de aprendizaje computacional más ampliamente usados para problemas de clasificación. Basan su funcionamiento en arreglos colectivos de neuronas artificiales donde cada una de éstas puede estar en uno de dos estados, encendida o apagada. Cada neurona tiene entradas que provienen del ambiente o pueden ser salidas de otras neuronas. Cada conexión (arco) entre neuronas tiene asociado un peso de conexión o sináptico, que es donde se codifica el aprendizaje de la red completa. El número de neuronas en la capa de entrada está en función de la dimensión del vector de características. La red neuronal puede tener una o más capas ocultas, así como una o varias neuronas en cada capa oculta. El número de neuronas de salida de la red dependerá de la forma de codificar el número de clases del problema.

La Fig. 2 muestra un ejemplo de una red neuronal para un problema de clasificación que recibe 2 elementos de entrada y tiene una capa de salida con 1 neurona.

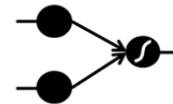


Figura 2. Modelo de red neuronal artificial de 3 neuronas.

La parte de la neurona que realiza el procesamiento es la función de transferencia, recibe como entrada la sumatoria ponderada de los arcos entrantes (pesos multiplicados por las salidas de neuronas correspondientes), y genera una salida. Entre las funciones de transferencia más populares se encuentran las lineales (para problemas de regresión) y las logísticas. La ecuación 1 muestra la función sigmoide, perteneciente a la familia de las logísticas.

$$h_{\theta}(x) = \frac{1}{1 + e^{-\theta^T x}} \quad (1)$$

donde  $x$  es el vector de entrada y  $\theta$  el vector de pesos de los arcos entrantes a la neurona. Las salidas de las neuronas se transmiten hacia adelante, siendo entradas a otras neuronas. Al final, en caso de ser aprendizaje supervisado, la salida de la red completa se compara con la etiqueta de los datos, si coinciden entonces el proceso se repite con el siguiente dato de entrenamiento, en caso contrario comienza un proceso de corrección de error llamado retropropagación. Durante este proceso de corrección, los pesos de la red se ajustan para minimizar el error de salida de la red neuronal.

##### 2) Regresión Logística

El modelo de regresión logística es otro algoritmo de clasificación. Se motiva a partir del deseo de modelar las probabilidades posteriores de las  $k$  clases a través de funciones lineales en el vector de entrada  $x$  [14]. Su funcionamiento se

basa en la optimización del vector  $\theta$  de parámetros de la ecuación 1. Una vez que se tienen estos parámetros, se asigna la clase con mayor probabilidad al vector de entrada, buscando con esto la clasificación de los datos.

#### IV. EXPERIMENTOS Y RESULTADOS

De los recorridos con los automóviles mencionados en la sección anterior se obtuvieron datos correspondientes a baches, boyas, topes, pavimento regular y pavimento irregular. Por pavimento irregular nos referimos a secciones de pavimento reparadas o deterioradas y que se encuentran a desnivel en referencia con el pavimento principal. El tipo y número de los eventos registrados se muestran en la tabla I.

TABLA I. CLASES DE RSDS DE INTERÉS.

| Evento           | Sentra | Tsuru | Total |
|------------------|--------|-------|-------|
| <b>Bache</b>     | 12     | 12    | 24    |
| <b>Boya</b>      | 14     | 32    | 46    |
| <b>Tope</b>      | 26     | 25    | 51    |
| <b>Regular</b>   | 33     | 27    | 60    |
| <b>Irregular</b> | 9      | 11    | 20    |

Es importante remarcar que el número de estos eventos es significativamente mayor que el dataset que usaron Fazeen et al. [12], e.g. Fazeen usaron 46 “bumps”, mientras que en este artículo entre boyas y topes se tienen 97 “bumps”. Otro punto a considerar es que en la literatura no existe ningún trabajo que discrimine entre boyas y topes, en este trabajo se presentan resultados que sugieren que esta clasificación es posible al considerar información de todos los ejes del acelerómetro.

La aplicación de algoritmos de aprendizaje computacional a datos extraídos de acelerómetros obedece en primera instancia a que existen patrones representativos que permiten realizar una clasificación para cada uno de RSDs. La Fig. 3 muestra las señales de casos representativos para cada uno de los RSDs. El panel A muestra la señal adquirida cuando se detecta una boya. Este tipo de evento genera cambios de energía similares a los “bumps” reportados por la literatura, con un incremento en el eje z seguido por un decremento en la magnitud de la señal. El panel B muestra la señal adquirida con un tipo de evento tope. Vemos que esta señal muestra muchas más perturbaciones que la señal de boya, sobre todo cuando consideramos los demás ejes. El panel C muestra la clase bache, aquí se puede observar un caso opuesto a la clase boya, primero aparece un decremento y después un incremento en el eje z, también existen perturbaciones en los demás ejes. El panel D muestra la clase camino regular, prácticamente aquí no se detecta la presencia de ningún evento de interés, sólo existe la lectura del acelerómetro con ruido del ambiente y perturbaciones despreciables. En cada una de las figuras del panel se puede observar la variación de los ejes correspondientes cuando se detecta el evento en función del tiempo. El reto para los algoritmos de clasificación es identificar los patrones que estas señales guardan y clasificar el RSD en base a un vector de características.

Podemos observar en estas gráficas que al identificarse un evento de interés los tres ejes son perturbados, es decir, existe información en los tres ejes que puede ser útil para realizar una clasificación de la información. Los procedimientos de umbral que se proponen en la literatura basan su identificación principalmente en un solo eje, el eje z, el cual se encuentra perpendicular al piso. La razón para tomar este eje como el único de referencia es que la principal perturbación ocurre de manera vertical en todos los eventos. No es sino hasta Fazeen et al. [12] que proponen el uso del eje x como auxiliar del eje z para lograr incrementar la efectividad en la identificación de baches.

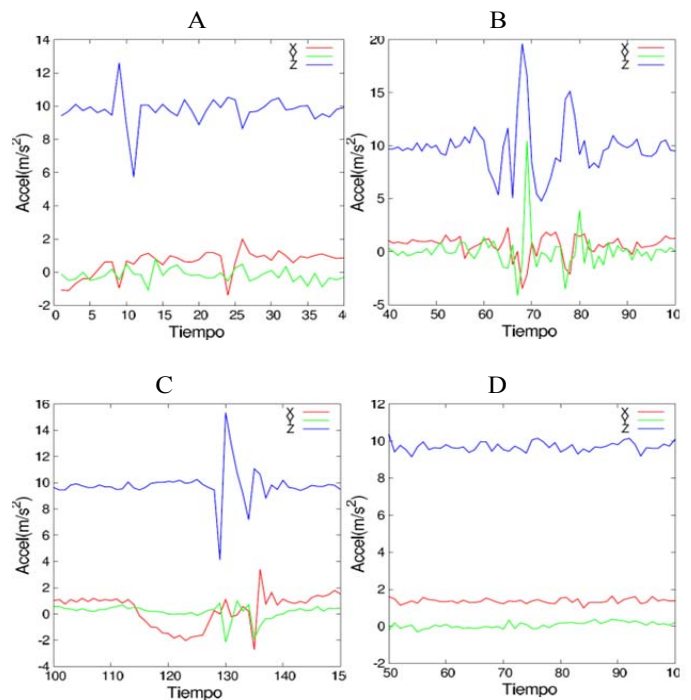


Figura 3. Patrones de señales de casos representativos para RSDs. A): Boya, B) tope, C) bache, D) camino regular.

Para la identificación de los eventos por medio de los algoritmos de aprendizaje, fue necesario proponer un vector de características que considerara información de variación de los 3 ejes en las lecturas del acelerómetro. Este vector de características considera los valores máximos y mínimos de los 3 ejes para cada uno de los eventos, por lo tanto la dimensión del vector es 6. La Fig. 4 muestra la identificación de los puntos que componen el vector de características para un evento de tipo tope.

Para la fase experimental se crearon varios datasets para contrastar la factibilidad de discriminación de los RSDs y para contrastar la potencia de los algoritmos de clasificación. Estos datasets los ponemos a disposición de la comunidad científica con la finalidad de motivar la discusión y establecer suelo fértil para elaborar comparativas de métodos y procedimientos para estos problemas. Se pueden consultar en <https://sites.google.com/site/luiscarlosgonzalezgurrola/dataset>.



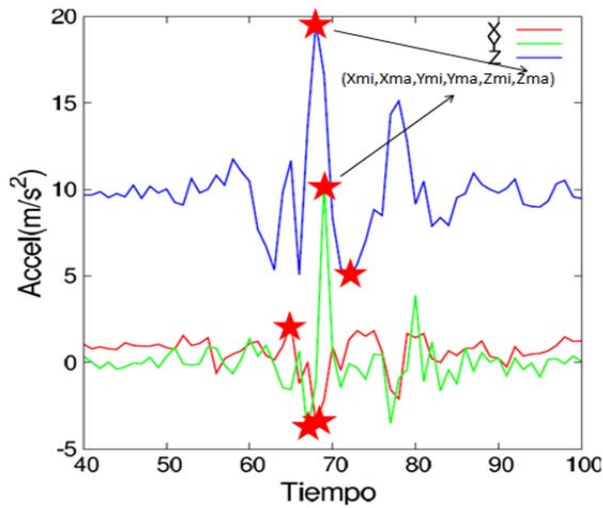


Figura 4. Vector de características a partir de una señal de evento. Las características son los valores máximos y mínimos de cada uno de los ejes.

Se compararon 2 algoritmos para la tarea de clasificación. El proceso de clasificación se desarrollo off-line. El primer algoritmo fue una red neuronal en código python con 1 capa oculta y 5 neuronas. El modelo y sus parámetros se eligieron de forma manual después de intentar varias configuraciones y observar que estos fueron los que mejor desempeño tuvieron. Se utilizó una tasa de aprendizaje de 0.1 y momentum de 0.9. Para evitar sobre entrenamiento los datos se dividieron en 50% para datos de entrenamiento, 25% para datos de validación y 25% para datos de prueba.

La primera tarea de clasificación consistió en observar si era posible discriminar la clase “tope” de la clase “boya”. De acuerdo a la revisión de literatura que realizamos, este tipo de experimento no se ha reportado, por lo que estos resultados son los primeros en publicarse. Con este propósito, generamos el dataset con nombre “twoBumps” que contiene 3 clases, la clase boya, la clase tope y la clase camino regular, después aplicamos la red neuronal. La tabla II muestra la matriz de confusión para este experimento.

La matriz de confusión para la clasificación de estas 3 clases, tiene un porcentaje de exactitud del 87.17% (34/38), lo cual sugiere que existe información que puede ser utilizada para considerar cada uno de estos eventos como clases diferentes. La matriz se interpreta de la forma siguiente. Todos los elementos de la diagonal principal son clasificaciones correctas, mientras que aquellos fuera de ésta son errores. Por ejemplo, la red neuronal identificó un falso positivo para la clase tope.

TABLA II. MATRIZ DE CONFUSIÓN PARA LA CLASIFICACIÓN DE 3 CLASES. LA CLASE “BUMPS” AHORA SE DIVIDE EN BOYA O TOPE Y TAMBIEN SE INCLUYE CAMINO REGULAR.

|         | Boya | Tope | Regular |
|---------|------|------|---------|
| Boya    | 10   | 2    | 0       |
| Tope    | 2    | 10   | 0       |
| Regular | 0    | 1    | 14      |

Para comparar las capacidades de discriminación de los algoritmos se crearon otros conjuntos de datos y se repitió el mismo proceso de clasificación desarrollado para el dataset “twoBumps”, ver tabla III.

TABLA III. DATASETS CREADOS PARA FASE EXPERIMENTAL.

| Nombre del dataset  | Descripción  |
|---------------------|--|
| <b>TwoBumps</b>     | Contiene la clase topes, boyas (junto con la clase camino regular).                          |
| <b>ThreeClasses</b> | Agrupar topes y boyas en la clase “bumps”, incluye baches y camino regular.                  |
| <b>FourClasses</b>  | Contiene topes, boyas, baches y camino regular.  |
| <b>FourClassesB</b> | Agrupar topes y boyas en la clase “bumps”, incluye baches camino regular y camino irregular. |
| <b>FiveClasses</b>  | Contiene topes, boyas, baches, camino regular y camino irregular                             |

Para contrastar la red neuronal, el segundo algoritmo que se utilizó fue la regresión logística con función de costo regularizada, la cual basa su funcionamiento en la optimización de parámetros de la ecuación 1. Este algoritmo se implementó en Matlab y se utilizaron funciones propias del programa para minimizar la función de error, e.g. fmincg. La regresión logística se aplicó al mismo conjunto de datos que la red neuronal. La tabla IV muestra la comparación en porcentaje de exactitud de ambos algoritmos para todos los conjuntos de datos que se utilizaron.

TABLA IV. COMPARACIÓN DE LA RED NEURONAL CON LA REGRESIÓN LOGÍSTICA PARA TODOS LOS DATASETS.

| Dataset             | Red neuronal  | Regresión logística |
|---------------------|---------------|---------------------|
| <b>TwoBumps</b>     | <b>87.17%</b> | 80.89%              |
| <b>ThreeClasses</b> | <b>91.1%</b>  | 86.18%              |
| <b>FourClasses</b>  | <b>77.7%</b>  | 69.06%              |
| <b>FourClassesB</b> | <b>86.27%</b> | 75.84%              |
| <b>FiveClasses</b>  | <b>76.47%</b> | 61.35%              |

## V. DISCUSIÓN

A pesar de ser un problema de clasificación bien identificado y planteado en la literatura es de llamar la atención que procedimientos de aprendizaje computacional no hayan sido usados de forma intensiva para lograr incrementar los porcentajes de éxito para la clasificación de eventos de irregularidades en caminos y carreteras. La mayoría de los procedimientos basan su funcionamiento en métodos heurísticos donde se mide el incremento de la señal del eje z con respecto a un umbral previamente definido. Se ha encontrado evidencia que los demás ejes del acelerómetro también sufren perturbaciones que bien caracterizadas pueden

ser utilizados para mejorar los porcentajes de clasificación. Un ejemplo que proponemos en este artículo es la clase “bumps”. De acuerdo a experimentos que se muestran, es posible lograr una clasificación mucho más fina de estos eventos en las clases topes y boyas, con un porcentaje de exactitud mayor al 87%. Esto abre la posibilidad de discriminar diferentes tipos de baches, de acuerdo a su tamaño, profundidad y características, lo cual permitiría eventualmente generar una lista de prioridades para su posible bacheo por la autoridad correspondiente.

La comparación más cercana que podemos hacer de los experimentos aquí reportados con experimento de la literatura es con Fazeen et al. [12], donde reportan un porcentaje de exactitud de su método de hasta un 85%. El dataset más parecido que se tiene con ellos es el de “FourClassesB”, este dataset tiene una clase menos, la clase “Uneven”, pero tiene muchos más eventos identificados (prácticamente el doble). El porcentaje de exactitud de la red de neuronal para clasificar este dataset es de 86.27%. Estos resultados son alentadores, ya que da indicios que utilizar información de todos los ejes, y no sólo del eje z, puede mejorar los porcentajes de exactitud en la clasificación. Esta información bien codificada en un vector de características en conjunción con procedimientos algorítmicos que históricamente han explotado patrones en los datos pudiera detonar en porcentajes de éxito no reportados hasta la fecha en la literatura. También se identificó que la clase más susceptible de error fue la clase bache, que en algunas ocasiones los algoritmos la confundían con la clase “bump”.

También llama la atención la falta de datasets públicos que permitan comparar y contrastar procedimientos de solución. Hasta la fecha, todos los autores han demostrado la utilidad de métodos en datasets propios, sin la capacidad real de contrastar alcances y limitaciones de los resultados. Es necesario abrir a la comunidad estos experimentos, que permitan avanzar en este problema de interés.

## VI. CONCLUSIONES.

En este trabajo se ha primeramente revisado las oportunidades que ofrecen las tecnologías emergentes para recolectar datos representativos de diversos tipos de irregularidades en caminos y carreteras, y en consecuencia contar con información que podría dar soporte a los departamentos que administran y mantienen la infraestructura vial en países en desarrollo.

Posteriormente se describieron las estrategias utilizadas en este proyecto para la recolección de datos asociados a RSDs utilizando dispositivos móviles, y la aplicación de algoritmos de aprendizaje computacional para la clasificación de baches, topes, boyas y pavimento irregular.

Utilizando como referencia el trabajo desarrollado por Fazeen et al. [12], se contrastan los resultados obtenidos y se observa un incremento en los porcentajes de clasificación para los RSDs. Este trabajo es uno de los primeros en contrastar algoritmos del área de aprendizaje de máquina, como lo son redes neuronales y regresión logística, para la clasificación de RSDs.

Los resultados obtenidos hasta el momento indican la

posibilidad de extender las técnicas de aprendizaje computacional utilizadas con el objetivo de incrementar los porcentajes de éxito en la clasificación de RSDs.

Un siguiente nivel de desarrollo para este trabajo incluye la extensión de las zonas de recolección de datos y el uso de otros tipos de automóviles para fortalecer el “ground truth” de los tipos de RSDs.

Adicionalmente, existe trabajo en desarrollo para explorar la implementación parcial o total del sistema de clasificación en los teléfonos inteligentes. Finalmente, estamos en proceso de construir el sistema de información que brindará soporte tanto a los ciudadanos como el departamento de administración de carreteras en la ciudad de Chihuahua, Chihuahua, México.

## REFERENCIAS

- [1] Friedewald, Michael, and Olivier Da Costa. "Science and technology roadmapping: Ambient intelligence in everyday life (Aml@ Life)." *Karlsruhe: Fraunhofer-Institut für System-und Innovationsforschung (FhG-ISI)* (2003).
- [2] A. Foth, Marcus. *Handbook of research on urban informatics: The practice and promise of the real-time city*. Information Science Reference, IGI Global, 2009.
- [3] Koch, Fernando, Carlos Cardonha, Jan Marcel Gentil, and Sergio Borger. "A platform for citizen sensing in sentient cities." In *Citizen in Sensor Networks*, pp. 57-66. Springer Berlin Heidelberg, 2013.
- [4] H. Zheng, Yu, Yanchi Liu, Jing Yuan, and Xing Xie. "Urban computing with taxicabs." In *Proceedings of the 13th international conference on Ubiquitous computing*, pp. 89-98. ACM, 2011.
- [5] Jengo, Christopher M., David Hughes, Joseph D. LaVeigne, and Ivan Curtis. "Pothole Detection and Road Condition Assessment using Hyperspectral Imagery." In *Proceedings of the American Society for Photogrammetry & Remote Sensing (ASPRS) 2005 Annual Conference*, pp. 7-11. 2005.
- [6] Kertész, I., T. Lovas, and A. Barsi. "Photogrammetric pavement detection system." *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 37 (2008).
- [7] Mednis, Artis, Girts Strazdins, Martins Liepins, AndrisGordjusins, and Leo Selavo. "Roadmic: Road surface monitoring using vehicular sensor networks with microphones." In *Networked Digital Technologies*, pp. 417-429. Springer Berlin Heidelberg, 2010.
- [8] De Silva, Girisha D., Ravin S. Perera, Nayanajith M. Laxman, Kenneth M. Thilakarathna, Chamath I. Keppitiyagama, and Kasun De Zoysa. "Automated Pothole Detection System."
- [9] Eriksson, Jakob, Lewis Girod, Bret Hull, Ryan Newton, Samuel Madden, and HariBalakrishnan. "The pothole patrol: using a mobile sensor network for road surface monitoring." In *Proceedings of the 6th international conference on Mobile systems, applications, and services*, pp. 29-39. ACM, 2008.
- [10] Astarita, Vittorio, Maria Vittoria Caruso, Guido Danieli, Demetrio Carmine Festa, Vincenzo Pasquale Giofrè, Teresa Iuele, and Rosolino Vaiana. "A mobile application for road surface quality control: UNIQALroad." *Procedia-Social and Behavioral Sciences* 54 (2012): 1135-1144.
- [11] Mohan, Prashanth, Venkata N. Padmanabhan, and RamachandranRamjee. "Nericell: rich monitoring of road and traffic conditions using mobile smartphones." In *Proceedings of the 6th ACM conference on Embedded network sensor systems*, pp. 323-336. ACM, 2008.
- [12] Fazeen, Mohamed, Brandon Gozick, Ram Dantu, MoizBhukhiya, and Marta C. González. "Safe driving using mobile phones." *Intelligent Transportation Systems, IEEE Transactions on* 13, no. 3 (2012): 1462-1468.
- [13] Alpaydin Ethem. *Introduction to Machine Learning*. The MIT Press, 2009.
- [14] Hastie Trevor, Tibshirani Robert, Jerome Friedman. *The Elements of Statistical Learning: Data mining, Inference and Prediction*. Springer, 2011.



**Luis Carlos González-Gurrola** received the Master degree in Computer Science from the Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California, México, in 2004, and the PhD in Information Technology from The University of North Carolina at Charlotte, Charlotte N.C, USA, in 2011. His current research interest are pattern recognition and combinatorial optimization problems.



**Fernando Martínez-Reyes** received the Master degree in Industrial Electronic Engineering from the Instituto Tecnológico de Chihuahua - Chihuahua, México, in 1997, and the PhD in Computer Science from The University of Nottingham, Nottingham, UK, in 2009. His current research interest are mobile computing, sensor networks and human-computer interactions.



**Manuel Ricardo Carlos-Loya** is currently a M.Eng. student in Software Engineering at the Universidad Autónoma de Chihuahua. He received a B.Eng. in Computer Systems Engineering from the Instituto Tecnológico de Chihuahua II, Chihuahua, México in 2010. His current research interest includes cloud computing and e-collaboration.