

Document downloaded from:

<http://hdl.handle.net/10251/38164>

This paper must be cited as:

Fogue, M.; Garrido, P.; Martínez, F.J.; Cano Escribá, J.C.; Tavares De Araujo Cesariny Calafate, C.M.; Manzoni, P. (2014). A system for automatic notification and severity estimation of automotive accidents. *IEEE Transactions on Mobile Computing*. 13(5):948-963. doi:10.1109/TMC.2013.35.



The final publication is available at

<http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6477047>

Copyright Institute of Electrical and Electronics Engineers (IEEE)

A System for Automatic Notification and Severity Estimation of Automotive Accidents

Manuel Fogue, Piedad Garrido, Francisco J. Martinez

University of Zaragoza, Spain

Email: {m.fogue, piedad, f.martinez}@unizar.es

Juan-Carlos Cano, Carlos T. Calafate, Pietro Manzoni

Universitat Politècnica de València, Spain

Email: {jucano, calafate, pmanzoni}@disca.upv.es

Abstract

New communication technologies integrated into modern vehicles offer an opportunity for better assistance to people injured in traffic accidents. Recent studies show how communication capabilities should be supported by artificial intelligence systems capable of automating many of the decisions to be taken by emergency services, thereby adapting the rescue resources to the severity of the accident and reducing assistance time. To improve the overall rescue process, a fast and accurate estimation of the severity of the accident represent a key point to help the emergency services to better estimate the required resources. This paper proposes a novel intelligent system which is able to automatically detect road accidents, notify them through vehicular networks, and estimate their severity based on the concept of data mining and knowledge inference. Our system considers the most relevant variables that can characterize the severity of the accidents (variables such as the vehicle speed, the type of vehicles involved, the impact speed, and the status of the airbag). Results show that a complete Knowledge Discovery in Databases (KDD) process, with an adequate selection of relevant features, allows generating estimation models able to predict the severity of new accidents. We develop a prototype of our system based on off-the-shelf devices, and validate it at the Applus+IDIADA Automotive Research Corporation facilities, showing that our system can notably reduce the time needed to alert and deploy the emergency services after an accident takes place.

Index Terms

KDD; Data Mining; Vehicular Networks; Traffic Accidents Assistance

I. INTRODUCTION

During the last decades, the total number of vehicles in our roads has experienced a remarkable growth, making traffic density higher and increasing the drivers' attention requirements. The immediate effect of this situation is the dramatic increase of traffic accidents on the road, representing

a serious problem in most countries. As an example, 2,478 people died in Spanish roads in 2010, which means one death for every 18,551 inhabitants [1], and 34,500 people in the whole European Union died as a result of a traffic accident in 2009 [2].

To reduce the number of road fatalities, vehicular networks will play an increasing role in the *Intelligent Transportation Systems* (ITS) area. Most ITS applications, such as road safety, fleet management, and navigation, will rely on data exchanged between the vehicle and the roadside infrastructure (V2I), or even directly between vehicles (V2V) [3]. The integration of sensing capabilities on-board of vehicles, along with peer-to-peer mobile communication among vehicles, forecast significant improvements in terms of safety in the near future.

Before arriving to the zero accident objective on the long term, a fast and efficient rescue operation during the hour following a traffic accident (the so-called *Golden Hour* [4]) significantly increases the probability of survival of the injured, and reduces the injury severity. Hence, to maximize the benefits of using communication systems between vehicles, the infrastructure should be supported by intelligent systems capable of estimating the severity of accidents, and automatically deploying the actions required, thereby reducing the time needed to assist injured passengers. Many of the manual decisions taken nowadays by emergency services are based on incomplete or inaccurate data, which may be replaced by automatic systems that adapt to the specific characteristics of each accident. A preliminary assessment of the severity of the accident will help emergency services to adapt the human and material resources to the conditions of the accident, with the consequent assistance quality improvement [5].

In this paper, we take advantage of the use of vehicular networks to collect precise information about road accidents that is then used to estimate the severity of the collision. We propose an estimation based on data mining classification algorithms, trained using historical data about previous accidents. Our proposal does not focus on directly reducing the number of accidents, but on improving post-collision assistance.

The rest of the paper is organized as follows: Section II presents the architecture of our proposed automatic system to improve accident assistance. Sections III, IV, and V provide details of our Knowledge Discovery in Databases (KDD) model adapted to the traffic accidents domain. Section VI presents the implemented prototype built to test our system and evaluates the obtained results of the validation process. Section VII reviews the related work on the improvement of traffic safety through telecommunication technologies, and data mining for accident severity estimation. Finally, Section VIII concludes this paper.

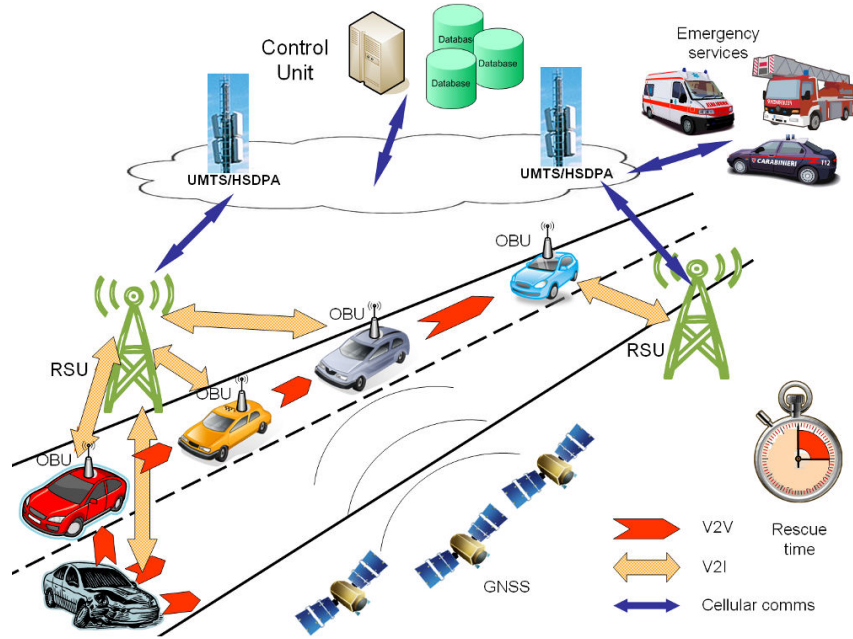


Fig. 1. Architecture of our proposed system for automatic accident notification and assistance using vehicular networks.

II. OUR PROPOSAL

Our approach collects information available when a traffic accident occurs, which is captured by sensors installed on-board the vehicles. The data collected are structured in a packet, and forwarded to a remote Control Unit through a combination of V2V and V2I wireless communication. Based on this information, our system directly estimates the accident severity by comparing the obtained data with information coming from previous accidents stored in a database. This information is of utmost importance, for example, to determine the most suitable set of resources in a rescue operation. Since we want to consider the information obtained just when the accident occurs, to estimate its severity immediately, we are limited by the data automatically retrievable, omitting other information, e.g., about the driver’s degree of attention, drowsiness, etc.

A. Architecture Overview

Figure 1 presents the overview of the vehicular architecture used to develop our system. The proposed system consists of several components with different functions. Firstly, vehicles should incorporate an On-Board unit (OBU) responsible for: (i) detecting when there has been a potentially dangerous impact for the occupants, (ii) collecting available information coming from sensors in the vehicle, and (iii) communicating the situation to a Control Unit (CU) that will accordingly address the handling of the warning notification. Next, the notification of the detected accidents is made through a combination of both V2V and V2I communications. Finally, the destination of all the collected

information is the Control Unit; it will handle the warning notification, estimating the severity of the accident, and communicating the incident to the appropriate emergency services.

The OBU definition is crucial for the proposed system. This device must be technically and economically feasible, as its adoption in a wide range of vehicles could become massive in a near future. In addition, this system should be open to future software updates. Although the design of the hardware to be included in vehicles initially consisted of special-purpose systems, this trend is heading towards general-purpose systems because of the constant inclusion of new services.

The information exchange between the OBUs and the CU is made through the Internet, either through other vehicles acting as Internet gateways (via UMTS, for example), or by reaching infrastructure units (Road-Side Units, RSU) that provide this service. If the vehicle does not get direct access to the CU on its own, it can generate messages to be broadcast by nearby vehicles until they reach one of the aforementioned communication paths. These messages, when disseminated among the vehicles in the area where the accident took place, also serve the purpose of alerting drivers traveling to the accident area about the state of the affected vehicle, and its possible interference on the normal traffic flow [6].

Our proposed architecture provides: (i) direct communication between the vehicles involved in the accident, (ii) automatic sending of a data file containing important information about the accident to the Control Unit, and (iii) a preliminary and automatic assessment of the damage of the vehicle and its occupants, based on the information coming from the involved vehicles, and a database of accident reports. According to the reported information and the preliminary accident estimation, the system will alert the required rescue resources to optimize the accident assistance.

B. On-Board Unit structure

The main objective of the proposed OBU lies in obtaining the available information from sensors inside the vehicle to determine when a dangerous situation occurs, and reporting that situation to the nearest Control Unit, as well as to other nearby vehicles that may be affected.

Figure 2 shows the OBU system, which relies on the interaction between sensors, the data acquisition unit, the processing unit, and wireless interfaces:

- In-vehicle sensors. They are required to detect accidents and provide information about its causes. Accessing the data from in-vehicle sensors is possible nowadays using the On-Board Diagnostics (OBD) standard interface [7], which serves as the entry point to the vehicle's internal bus. This standard is mandatory in Europe and USA since 2001. This encompasses the majority of the vehicles of the current automotive park, since the percentage of compatible vehicles will keep growing as very old vehicles are replaced by new ones.

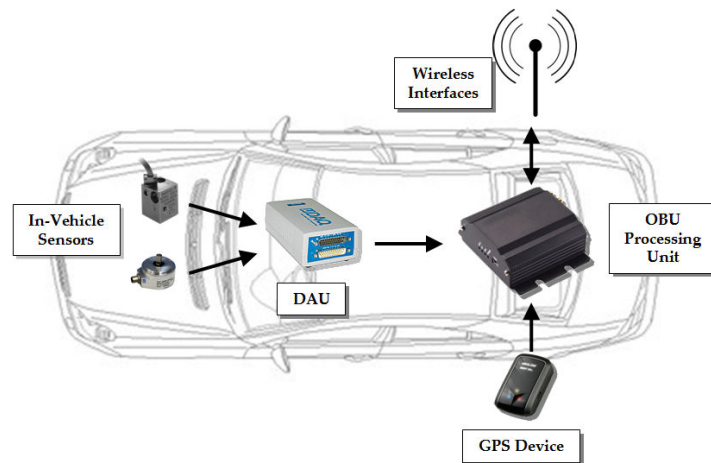


Fig. 2. On-Board Unit structure diagram.

- Data Acquisition Unit (DAU). This device is responsible for periodically collecting data from the sensors available in the vehicle (airbag triggers, speed, fuel levels, etc.), converting them to a common format, and providing the collected data set to the OBU Processing Unit.
- OBU Processing Unit. It is in charge of processing the data coming from sensors, determining whether an accident occurred, and notifying dangerous situations to nearby vehicles, or directly to the Control Unit. The information from the DAU is gathered, interpreted and used to determine the vehicle's current status. This unit must also have access to a positioning device (such as a GPS receiver), and to different wireless interfaces, thereby enabling communication between the vehicle and the remote control center.

C. Control Unit structure

The Control Unit (CU) is associated to the response center in charge of receiving notifications of accidents from the OBUs installed in vehicles. In particular, the Control Unit is responsible for dealing with warning messages, retrieving information from them, and notifying the emergency services about the conditions under which the accident occurred.

Figure 3 shows the modules included in the Control Unit to achieve all its objectives within our proposed system:

- Reception/interpretation module. The first step for the CU is to receive a warning message from a collided vehicle, and so there must be a module waiting for the arrival of messages and retrieving the values from the different fields.
- Accident severity estimation module. When a new accident notification is received, this module will determine how serious the collision was, and the severity of the passengers' injuries.

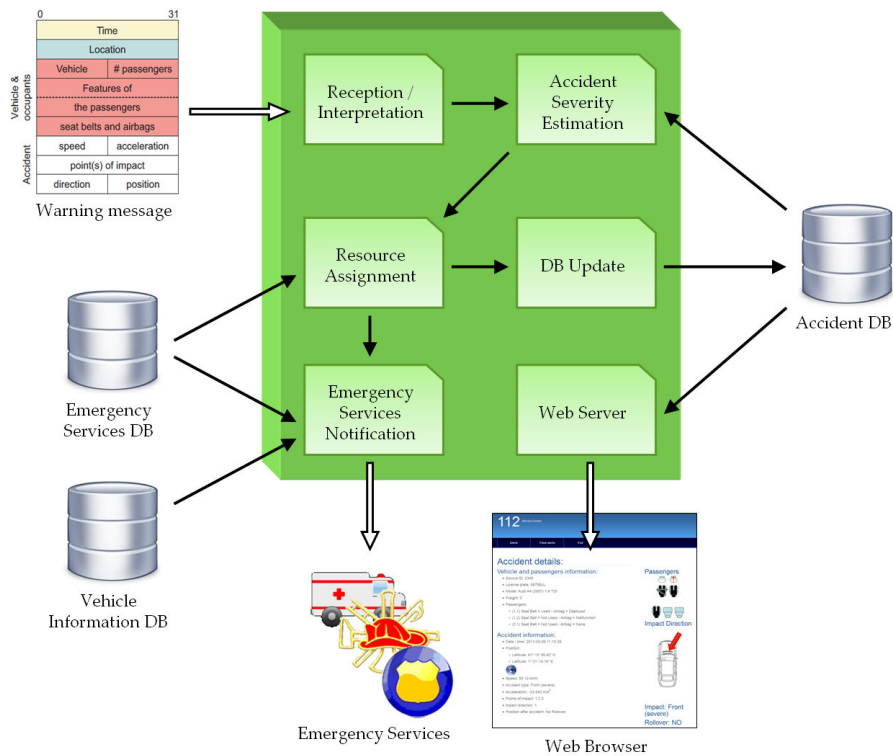


Fig. 3. Control Unit modular structure.

- Resource assignment module. After deciding the severity of the accident, an additional module is used to define resource sets adapted to the specific situation.
- Database update module. The data collected from the notified accident are stored into the existing database of previous accidents, increasing the knowledge about the accident domain.
- Web Server module. The Control Unit incorporates a Web Server to allow easy visualization of the historical information recorded and the current accident situations requiring assistance. A web interface was chosen in order to increase user friendliness and interoperability.
- Emergency services notification module. When the information has been correctly managed, the notification module sends messages to the emergency services including all the information collected, the estimated severity, the recommended set of resources, as well as additional information about the vehicles involved in the collision (for preliminary planning of the rescue operation). The information about vehicles consists of standard rescue sheets, which highlight the important or dangerous parts of a specific vehicle that should be taken into account during a rescue operation: batteries, fuel tanks, etc., as shown in Figure 4.

One of the most important modules in the Control Unit is in charge of the Accident Severity Estimation, i.e., providing a relative measure of the potential effect of the collision on the integrity of the vehicles and people involved. To obtain this estimation, we make use of historical information

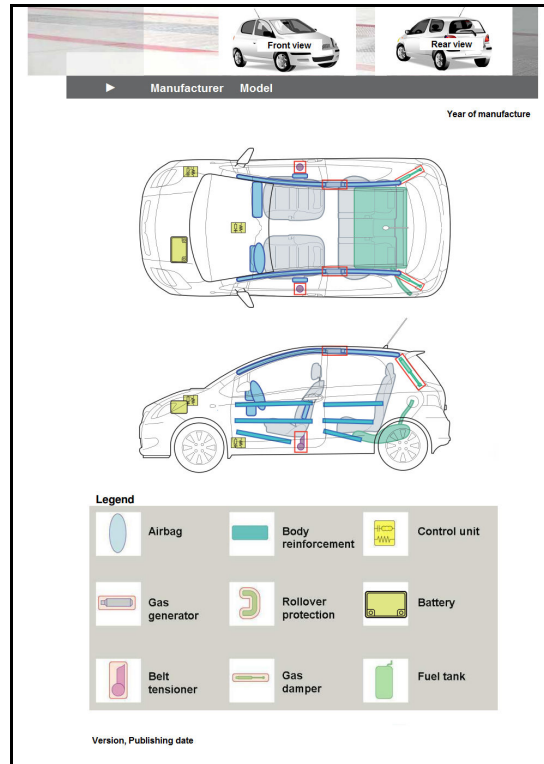


Fig. 4. Example of standard rescue sheet.

about previous accidents contained in an existing database, through a process of Knowledge Discovery in Databases (KDD).

The KDD approach can be defined as the nontrivial process of identifying valid, novel, potentially useful, and understandable patterns from existing data [8]. The KDD process begins with the understanding of the application specific domain and the necessary prior knowledge. After the acquisition of initial data, a series of phases are performed:

- 1) Selection: This phase determines the information sources that may be useful, and then it transforms the data into a common format.
- 2) Preprocessing: In this stage, the selected data must be cleaned (noise reduction or modeling) and preprocessed (missing data handling).
- 3) Transformation: This phase is in charge of performing a reduction and projection of the data to find relevant features that represent the data depending on the purpose of the task.
- 4) Data mining: This phase basically selects mining algorithms and selection methods which will be used to find patterns in data.
- 5) Interpretation/Evaluation: Finally, the extracted patterns must be interpreted. This step may also include displaying the patterns and models, or displaying the data taking into account such models.

We propose to develop a complete KDD process, starting by selecting a useful data source containing instances of previous accidents. The data collected will be structured and preprocessed to ease the work to be done in the transformation and data mining phases. The final step will consist on interpreting the results, and assessing their utility for the specific task of estimating the severity of road accidents. The phases from the KDD process will be performed using the open-source Weka collection, which is a set of machine learning algorithms [9]. Weka is open source software issued under the GNU General Public License which contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. We will deal with road accidents in two dimensions: (i) damage on the vehicle (indicating the possibility of traffic problems or the need of cranes in the area of the accident), and (ii) passenger injuries. These two dimensions seem to be related, since heavily damaged vehicles are usually associated with low survival possibilities of the occupants. Consequently, we will use the estimations obtained with our system about the damage on the vehicle to help in the prediction of the occupants' injuries.

Finally, our system will benefit from additional knowledge to improve its accuracy, grouping accidents according to their degree of similarity. We can use the criteria used in numerous studies about accidents, including some tests such as the Euro NCAP [10], in which crashes are divided and analyzed separately depending on the main direction of the impact registered due to the collision. The following sections contain the results of the different phases of our KDD proposal.

III. DATA ACQUISITION, SELECTION AND PREPROCESSING PHASES

Developing a useful algorithm to estimate accident severity needs historical data to ensure that the criteria used are suitable and realistic. The *National Highway Traffic Safety Administration* (NHTSA) maintains a database with information about road accidents which began operating in 1988: the General Estimates System (GES) [11]. The data for this database is obtained from a sample of real Police Accident Reports (PARs) collected all over the USA roads, and it is made public as electronic data sets [12].

In the traffic accidents domain, the most relevant sets of information in GES are: (i) *Accident*, which contains the crash characteristics and environmental conditions at the time of the accident, (ii) *Vehicle*, which refers to vehicles and drivers involved in the crash, and (iii) *Person*, i.e., people involved in the crash. We will integrate the data harvested during the year 2011 into two different self-built sets: one for the vehicles and another one for the occupants.

Using the data contained in the GES database, we classify the damage in vehicles in three categories: (i) *minor* (the vehicle can be driven safely after the accident), (ii) *moderate* (the vehicle shows defects that make it dangerous to be driven), and (iii) *severe* (the vehicle cannot be driven at all, and needs to be towed). Focusing on passenger injuries, we will also use three different classes to determine

their severity level: (i) *no injury* (unharmd passenger), (ii) *non-incapacitating injury* (the person has minor injuries that does not make him lose consciousness, or prevent him from walking), and (iii) *incapacitating or fatal injury* (the occupants' wounds impede them from moving, or they are fatal).

After preprocessing the selected GES data, no noise or inaccuracies were detected as all the nominal and numerical values contained reasonable values. Due to the large number of records available in the database, we decided to only use those accident records with all the required information complete. After removing incomplete instances, our data sets consist of 14,227 full instances of accident reports (5,604 front crashes, 4,551 side crashes, and 4,072 rear-end crashes). These accidents represent different types of collisions in both urban and inter-urban areas. The distribution of accidents depending on the area is the following:

- Front collisions: 1,418 (25.3%) in urban area, and 4,186 (74.7%) in inter-urban area.
- Side collisions: 1,593 (35.0%) in urban area, and 2,958 (65.0%) in inter-urban area.
- Rear-end collisions: 1,613 (39.6%) in urban area, and 2,459 (60.4%) in inter-urban area.

Using these data, we achieve a total distribution of 4,624 accidents in urban areas, which correspond to 32.5% of the total accidents, and 9,603 accidents in inter-urban areas (67.5%).

IV. TRANSFORMATION PHASE

This phase consists on developing a reduction and projection of the data to find relevant features that represent the characteristics of the data depending on the objective. We selected a potential subset of variables which could be obtained from the on-board sensors of the vehicle or auxiliary devices such as the GPS [13]. Those variables include the type of vehicle, the speed just before the accident, and the airbag status. **Even if the GES database does not include information about the the measured accelerations, this information could be filled using our proposed system with data collected from notified accidents, and incorporated in future versions of the classification algorithm.** Concerning passengers, there are specific characteristics for each person that are not directly accessible, but might help to improve the prediction accuracy. We added two of these personal variables to our data -age and sex-, which will be used to study their relevance on the injuries suffered. Weka provides a wide variety of feature selection algorithms. Among them, we selected three of the most commonly used:

- Correlation-based Feature Selection (*CfsSubsetEval*) [14]: This filtering algorithm considers that the appropriate subsets of attributes are those that contains features highly correlated with the class (accident severity), but not correlated with each other.
- Information gain selection (*InfoGainAttributeEval*): This metric is based on the change of entropy (measure of the uncertainty associated with a random variable) of a set of training examples when introducing knowledge about the values of a variable.

TABLE I
MOST RELEVANT VARIABLES FOR VEHICLE DAMAGE AND PASSENGER INJURY ESTIMATION.

Attribute	Vehicle damage				Passenger Injury			
	Full Set	Front	Side	Rear-end	Full Set	Front	Side	Rear-end
Body Type	✓		✓	✓	✓	✓	✓	✓
Light Condition	✓	✓		✓				✓
Model Year				✓				
Point of Impact	✓		✓		✓			
Road Align ^a						✓		
Road Profile ^b		✓				✓		
Rollover	✓	✓	✓	✓	✓		✓	✓
Speed	✓	✓	✓	✓	✓	✓	✓	
Speed limit	✓	✓	✓	✓	✓		✓	✓
Surface condition				✓				
Trailer			✓	✓	✓		✓	✓
Vehicle role ^c		✓	✓					✓
Weather			✓	✓				
Airbag					✓	✓	✓	✓
Age								
Restraint system					✓	✓	✓	
Seat position							✓	
Sex							✓	
Veh. damage estim.					✓	✓	✓	✓

^a Roadway alignment just prior to the vehicle's critical precrash event (straight, curve)

^b Roadway profile just prior to the vehicle's critical precrash event (level, grade, hillcrest, sag)

^c Determines whether the vehicle was the striking or the struck one

- Wrapper technique (*WrapperSubsetEval*) [15]: This method searches the entire space of possible variable subsets, aimed at finding the state which presents the highest accuracy. The wrapper approach normally requires more computing time, but it usually improves accuracy.

We determined the optimal variable subset with the three different schemes, and we chose for our final subset those variables selected by, at least, two of the previous algorithms. All the tested variables and the results of the feature selection process appear in Table I. The top part of the table contains variables about the vehicle involved in the accident, and hence also applicable to the occupants of that vehicle. The bottom part shows variables only applicable to individual occupants. We compared the results of the process when using the whole data set available (*Full Set*), and dividing these data into three subsets depending on the direction of the impact.

As shown, we find noticeable differences between the sets determined for the full set of accident data, and for each of the divisions depending on the direction of the impact. The most relevant

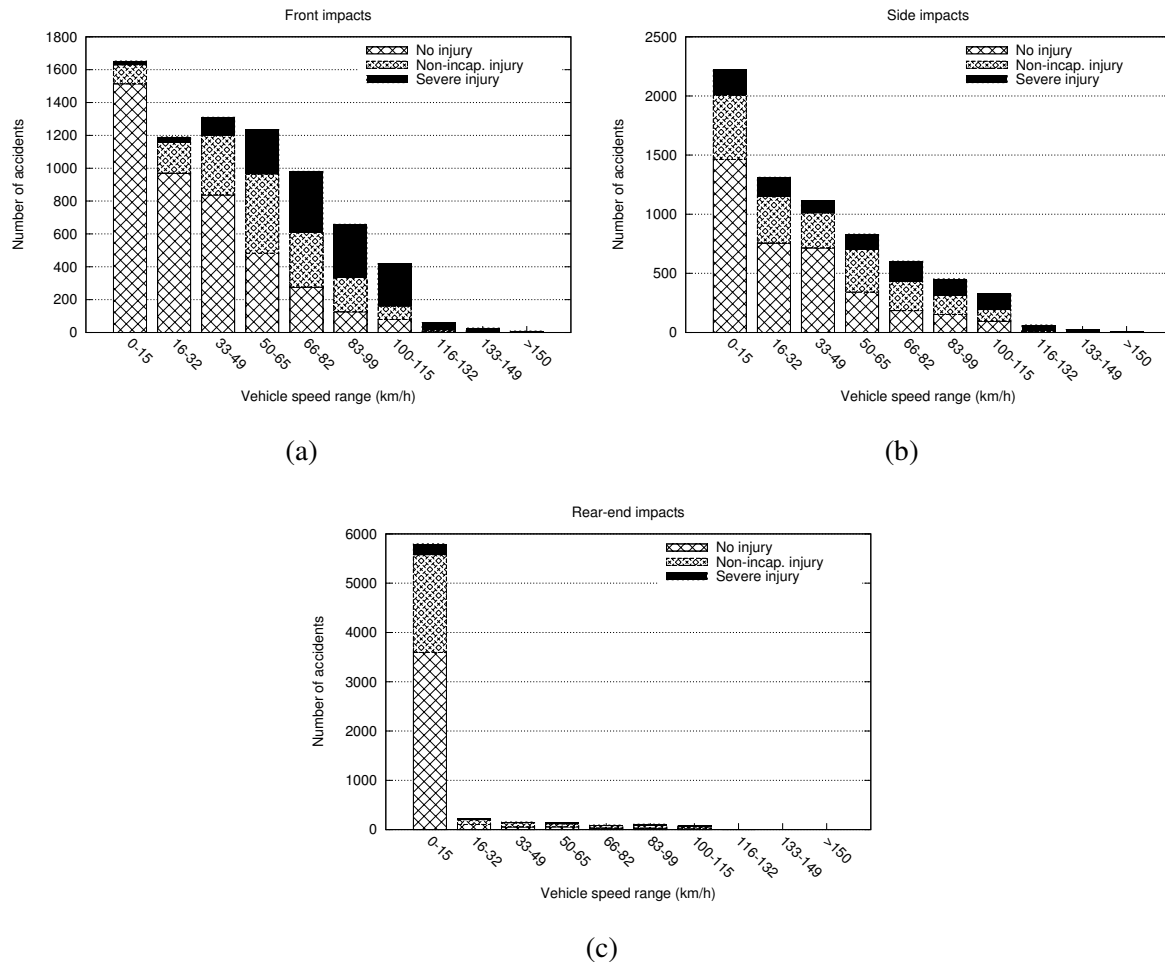


Fig. 5. Influence of the speed of the vehicle on the distribution of the severity of the passengers' injuries in (a) front, (b) side, and (c) rear-end impacts.

attributes (in almost all cases) are the body type, the occurrence of rollover, the speed, the speed limit, the presence of a trailer, the airbag status, and the estimation on vehicle damage. When we divide the instances, new important values appear, like the road profile when considering front accidents, or the light condition for rear-end crashes, which were not detected when using the full data set.

For accidents in which the vehicle strikes another urban infrastructure of the road, the attribute that offers a better approximation to determine the accident severity is the speed of the vehicle before the impact. Nevertheless, if you want to estimate the damage when the vehicle is struck by another one, the type of vehicle and the speed limit in the area are more important variables than the speed itself, as deduced by studying side and rear-end accidents. The speed limit in the area of the accident is a good estimator of the speed of other vehicles, and usually accidents occurring in highways are more severe than those in residential areas where the speed limit is lower.

This effect is shown in Figures 5 and 6, which highlight the influence of the speed of the vehicle and the speed limit on the severity of the passengers' injuries, respectively. The severity of front collisions

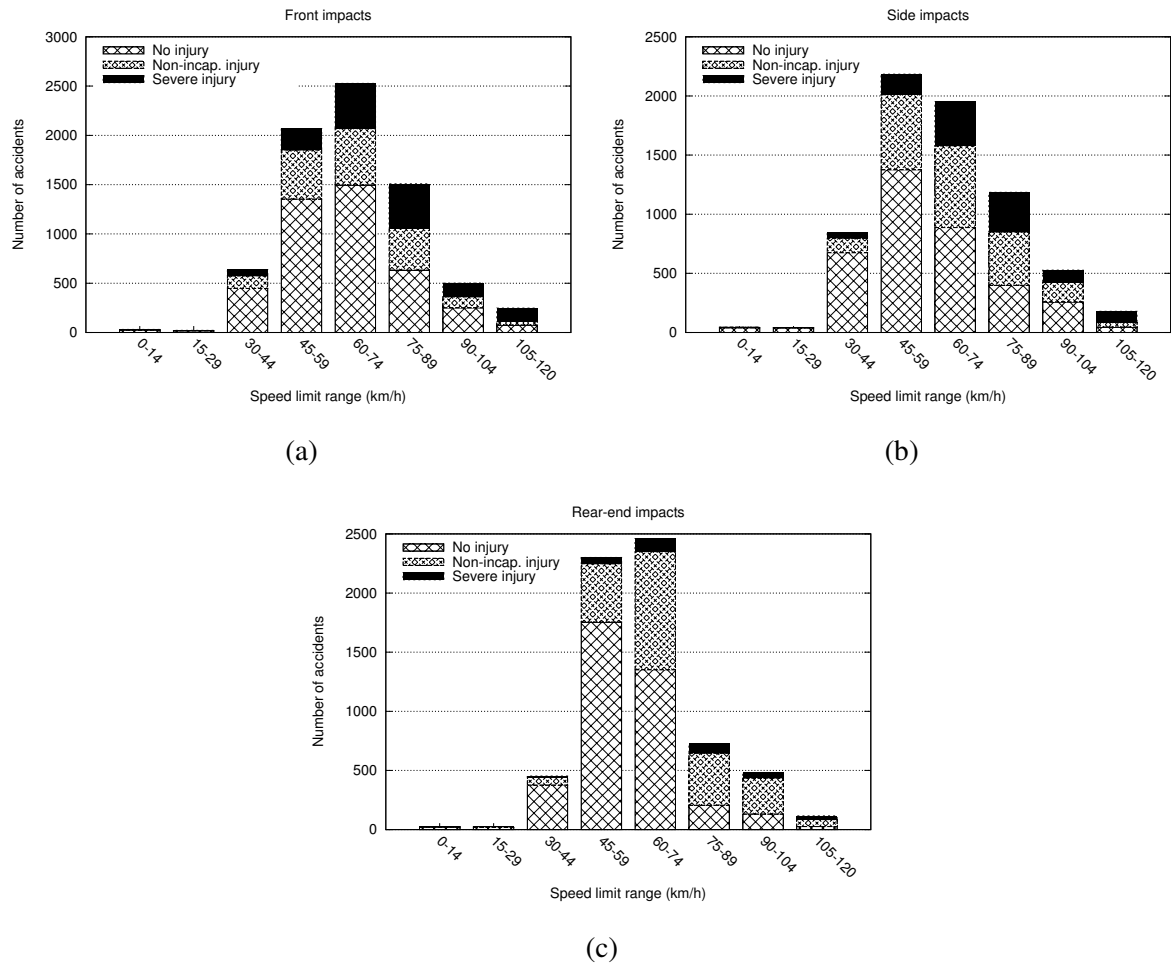


Fig. 6. Influence of the speed limit on the distribution of the severity of the passengers' injuries in (a) front, (b) side, and (c) rear-end impacts.

is clearly dependent on the speed of the vehicle itself, since more than half of the registered accidents occurred at speeds greater than 80 km/h resulted on severe injuries to the passengers. However, this dependence is less visible for side impacts, where the proportion between injury levels remains more stable, and especially for rear-end impacts, where almost 85% of the accidents happened when the vehicle was moving at speeds below 15 km/h. Using the speed limit as an indicator, Figure 6(c) shows that minor accidents are the dominant type for speed limits under 60 km/h, and that collisions above this speed were associated with different levels of severity on the passengers' injuries.

The selection of most of these variables is quite intuitive, but there are some important and non obvious considerations. The restraint system is not selected in all cases, as it would be reasonable, since it is not usually critical for the security of the passengers in rear-end collisions. Also, the year when the vehicle involved in the crash was manufactured is only relevant to determine the damages to the vehicle in rear-end collision. The presence of a trailer is also important, especially when the vehicle is struck by another one, which can cause dangerous situations such as uncontrolled

articulation or *jackknife*.

V. DATA MINING AND INTERPRETATION/EVALUATION PHASES

The most adequate data mining task for our interests is classification. Each instance has a record indicating its class membership, while the rest of the available attributes are used to predict the class of new instances. We selected three of the classification algorithms provided by Weka to study which one obtains the best results in terms of prediction accuracy:

- Decision trees: *J48* is an open source Java implementation of Quinlan's C4.5 algorithm [16] found in Weka, designed to build decision trees from a training data set. At each node of the tree, this algorithm selects the attribute that most effectively separates the sample set into subsets with a predominant class value. Most existing works in this area makes use of this tool to classify road accidents, and thus we will use it as well as reference.
- Bayesian networks: Bayesian networks allow modeling a phenomenon with a set of random variables and the dependency relationships between them. Those variables represent qualitative knowledge of the model by a directed acyclic graph in which the variables are represented as nodes, and the relationships of dependence and conditional independence between them are shown as arcs between the nodes. They also allow to express the strength of the relationship through probability distributions. We will use the *BayesNet* implementation of Bayesian networks, available in Weka, along with the K2 algorithm [17] to find the graph that best represents the set of dependence or independence in the data. This model has been rarely used in this domain, despite its power to represent complex phenomena.
- Support Vector Machines: SVMs are a set of supervised learning algorithms that are based on the construction of a set of hyperplanes in a high dimensional space (generated by a kernel function) to separate instances from different classes. We will use the SMO algorithm [18] to train the SVMs from the GES database.

These algorithms present different parameters that must be tuned to maximize the accuracy of the built models. The specific parameters used for each algorithm are: the prune level in decision trees, the number of parents in Bayesian networks, and the specific kernel function used in SVMs. We carried out several tests to obtain the sets of values that produce the best performance for the selected metrics, corresponding to the results shown.

The effectiveness of the classification can be measured using different metrics. One of the most used ones is the True Positive Rate (TP Rate) or percentage of instances correctly classified by the algorithm. Nevertheless, this metric could lead to wrong impressions, since those classifiers which focus their attention on the most frequent class would get a good value on this metric, but their utility is low as they are not able to differentiate among the existing classes properly. The False Positive Rate

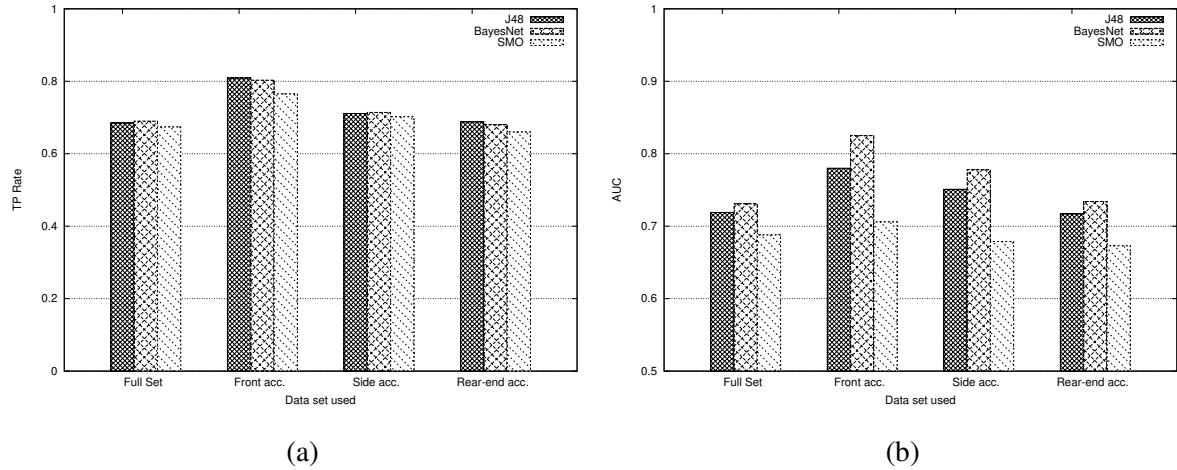


Fig. 7. Comparison of different data mining classification algorithms in the estimation of the damage on the vehicle due to the accident: (a) using the TP Rate metric, and (b) using the AUC metric.

(FP Rate), i.e., the percentage of classification errors with respect to the total number of instances belonging to the right class, should also be considered to determine the efficiency of the classification process.

To cope with the deficiencies of the TP Rate metric, we will also use the area under the ROC (Receiver Operating Characteristic) curve, abbreviated as AUC [19]. The possible values of this metric vary between 0.5 (for random classifiers with low efficiency) and 1.0 (for a perfect classifier), and it is computed using both the True Positive Rate and the False Positive (FP) Rate. Hence, the AUC metric presents some desirable features when compared to the overall accuracy [20]: (i) it provides an increased sensitivity in Analysis of Variance (ANOVA) tests because the standard error decreases as both AUC and the number of test samples increase, (ii) it is invariant to *a priori* class probabilities, and (iii) it offers an indication of the amount of work done by a classification scheme, providing low scores to both random and “one class only” classifiers.

A. Results of the classification

Figures 7 and 8 show the results of the selected algorithms for both the TP Rate and the AUC metrics. Results were obtained by using 10-fold cross validation, which reduces the dependence of the result from the classification process in terms of the partition made for training and validation.

When estimating the damages in vehicles (see Figure 7), the three algorithms showed similar performance using the TP Rate metric (though *SMO* is slightly worst in all cases), with an overall accuracy about 70 or 80%. However, there are noticeable differences between the schemes under the AUC metric, showing a clear advantage for the *BayesNet* algorithm. This means that Bayesian networks are more robust when facing doubtful cases, and they are not so focused in the majority

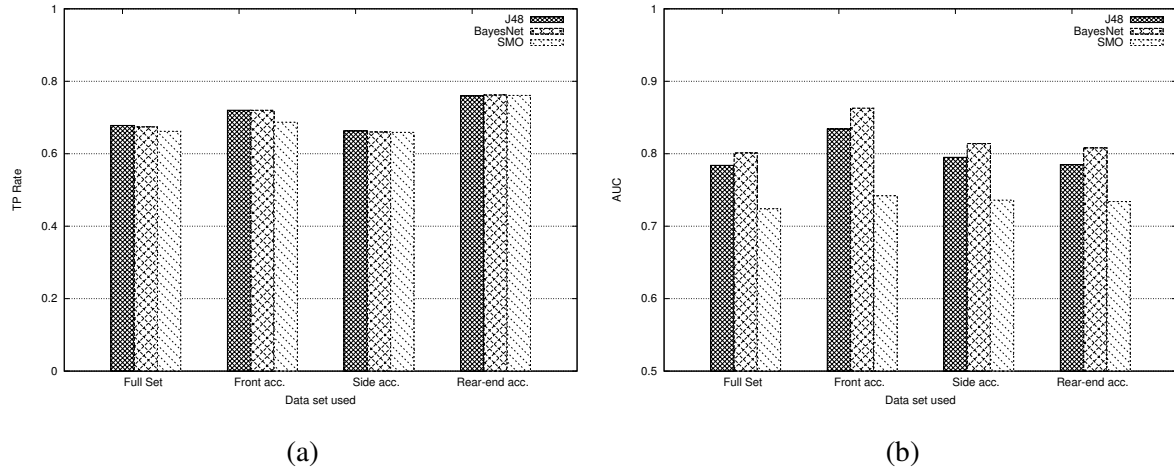


Fig. 8. Comparison of different data mining classification algorithms in the estimation of the injuries of the passengers in the vehicle: (a) using the TP Rate metric, and (b) using the AUC metric.

class. When we divide the accidents depending on the direction of the impact, we obtain a relevant increase on the accuracy for both metrics, showing average results much higher than those achieved with the full data set. Rear-end impacts were the most difficult to estimate, since there was a high proportion of instances where the car itself was struck by another vehicle, making it harder to estimate the damage without knowing all the details of the other vehicle.

If we try to estimate the injuries on passengers (see Figure 8), we observe a very similar trend. All the algorithms are very close in terms of the TP Rate metric (even closer than estimating the damage in the vehicle), and again the overall accuracy ranges from 70% to 80%. The differences between the three algorithms increase when we select the AUC metric, and *BayesNet* outperforms the other algorithms as well. Dividing the accident data in subsets also considerably improves the results under both metrics. However, when studying rear-end collisions, we obtain high values of TP Rate but lower results of AUC. This effect is due to the proportion of classes of this subset, since there are very few cases where passengers suffered from very severe injuries. In those cases, the algorithms reduce the probability of belonging to this class, and, even if the number of instances correctly classified is high, the AUC metric penalizes any excessive focus on a subset of the classes. This situation might improve if we were provided with additional information about the colliding vehicle, which will be possible using vehicular networks, including both V2V and V2I.

Since ROC analysis provides more information than using only the TP Rate metric, we will show some of the curves produced using the input data. ROC curves are commonly employed in two-class problems to represent the tradeoff between errors (false positives) and benefits (true positives) of a classifier. The resulting performance can be graphed in two dimensions for a good visualization. However, increasing the number of classes requires an increase of the number of dimensions of the

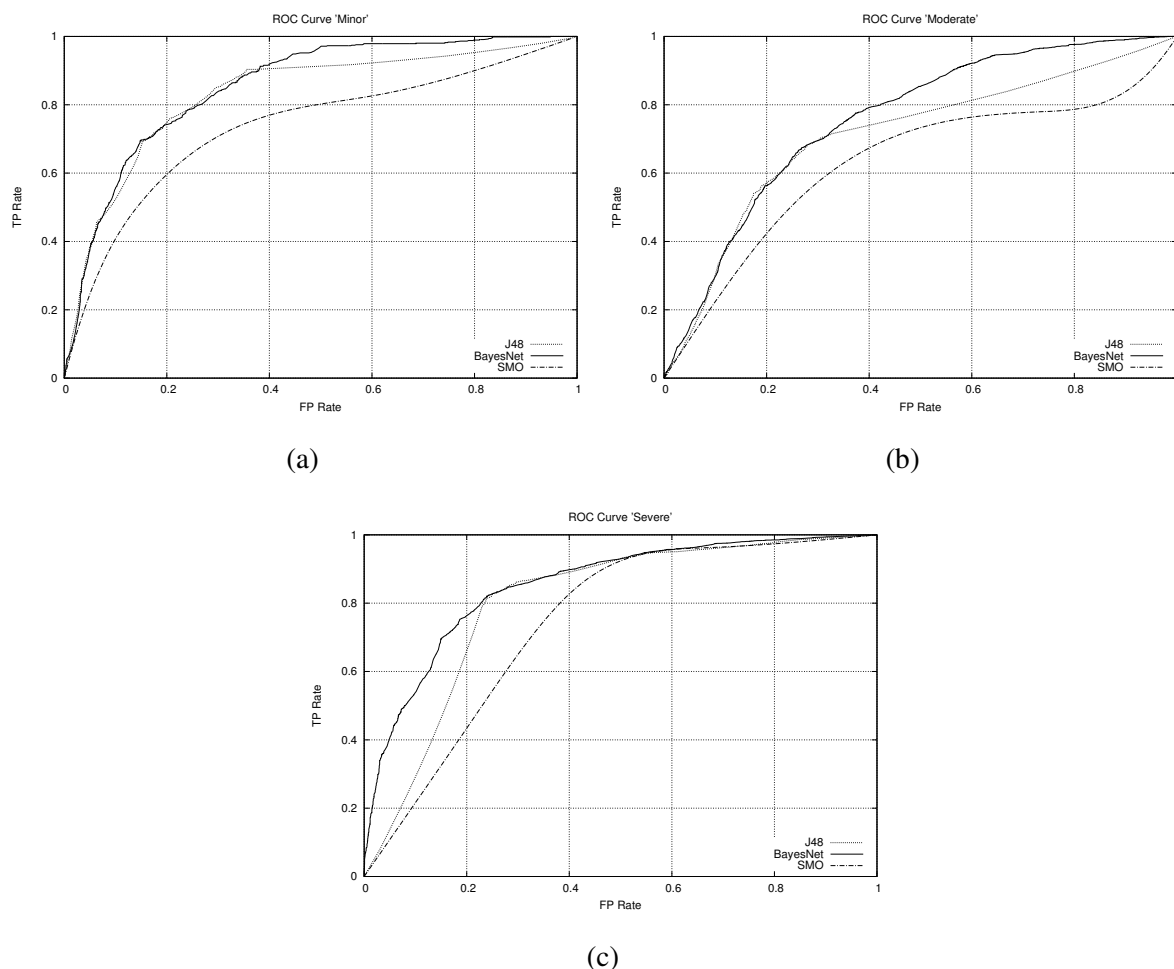


Fig. 9. Comparison of ROC curves using different data mining classification algorithms estimating the damage on the vehicle in a front collision, using as the positive value: (a) minor damage, (b) moderate damage, and (c) severe damage.

graph, thereby reducing visibility. A possible solution to this problem consists on producing a different ROC curve for each class [19], considering that the selected class represents the positive value, and the remaining classes represent negative values.

Figures 9 and 10 contain the ROC curves obtained for the three selected classification algorithms when studying front crashes, and their effect on both the damage on the vehicle and the passenger injuries. As shown, the tested classifiers achieve better results when discriminating the extreme values of the class, reducing their performance for intermediate values (*moderate damage* and *non-incapacitating injury*). There is a slight advantage for the Bayesian classifier in all the curves with respect to the classifier based on decision trees. The SMO algorithm is the less effective when attempting to obtain a good tradeoff between true and false positives, and it is clearly outperformed by the other evaluated algorithms.

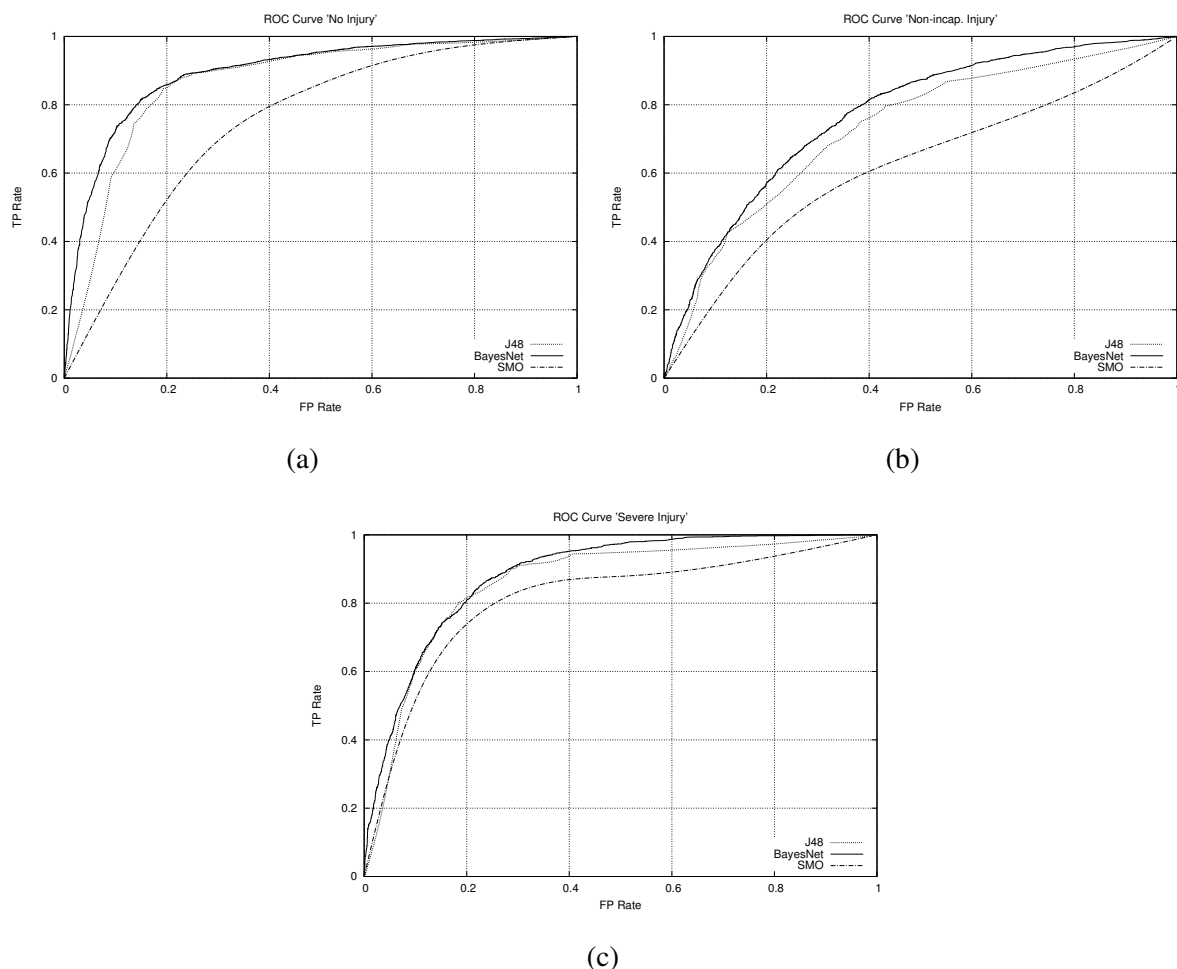


Fig. 10. Comparison of ROC curves using different data mining classification algorithms estimating the injuries of the passengers in a front collision, using as the positive value: (a) no injury, (b) non-incapacitating injury, and (c) incapacitating or fatal injury.

B. Bayesian models for accident severity estimation

Tables II and III show the conditional dependences found on the variables selected when estimating the severity of a road accident. In the Bayesian models, all the variables have at least one parent: the value of the class (severity of the damage on the vehicles or the passengers' injuries, respectively).

By studying the tables, we can find some intuitive connections, present in almost all cases. The type of vehicle (body type) presents strong connections with the presence of trailer, since more than 80% of the instances where the vehicle was towing a trailer, the vehicle was a heavy truck. The speed of the vehicle is the variable presenting more connections with other variables: speed limit (the speed is usually higher as the speed limit increases, representing an upper limit to the speed in most cases), rollover (only 15% of the registered overturned vehicles occurred at speeds under 60 km/h), light condition (vehicles usually reach the highest speed under good visibility conditions), and so on. Seat belts and airbags are found more often in touring cars; thus, the airbag status and the restraint

TABLE II

MAIN CONDITIONAL DEPENDENCES BETWEEN VARIABLES USED TO ESTIMATE THE DAMAGE ON THE VEHICLES.

Accident type	Dependences
Front accident	<i>Light condition, Speed limit</i> \rightarrow <i>Speed</i>
	<i>Speed</i> \rightarrow <i>Rollover, Vehicle role</i>
Side accident	<i>Body type</i> \rightarrow <i>Speed, Trailer, Rollover</i>
	<i>Speed limit</i> \rightarrow <i>Speed</i>
	<i>Speed</i> \rightarrow <i>Rollover, Vehicle role</i>
Rear-end accident	<i>Light condition, Surface condition, Speed limit</i> \rightarrow <i>Speed</i>
	<i>Speed</i> \rightarrow <i>Rollover</i>
	<i>Body type</i> \rightarrow <i>Trailer</i>
	<i>Weather</i> \rightarrow <i>Surface condition</i>

TABLE III

MAIN CONDITIONAL DEPENDENCES BETWEEN VARIABLES USED TO ESTIMATE THE INJURIES OF THE PASSENGERS.

Accident type	Dependences
Front accident	<i>Speed</i> \rightarrow <i>Restraint system, Vehicle damage</i>
	<i>Body type</i> \rightarrow <i>Speed, Airbag</i>
Side accident	<i>Speed limit</i> \rightarrow <i>Speed, Restraint system, Vehicle damage</i>
	<i>Speed</i> \rightarrow <i>Rollover, Vehicle damage</i>
	<i>Body type</i> \rightarrow <i>Airbag, Restraint system, Trailer, Seat position</i>
Rear-end accident	<i>Speed limit</i> \rightarrow <i>Vehicle damage, Vehicle role, Light condition, Body type</i>
	<i>Body type</i> \rightarrow <i>Airbag, Rollover</i>

systems usage is strongly related to the type of vehicle.

However, we find that some of the probabilities generated are not so intuitive. For example, the probability tables associated to front and rear-end impacts for passengers' injuries estimation show that, under daylight conditions, the majority of accidents are not very severe, while at dawn and dusk the probability of severe accidents increases up to 60%. This means that, when the light begins to decline but it is still not enough to use car lighting systems, accidents tend to be more severe.

The airbag status becomes the most determinant value for determining the injury severity on the passengers in front and side collisions. When the airbag does not need to be deployed, the impact is not usually dangerous for the passengers, whereas strong collisions where the airbags have to be deployed present a greater magnitude that could affect the passengers' health. To prove this observation, about 70% of the registered cases where the airbag was not deployed, resulted in few injuries on the passengers, and the occupants suffered from severe injuries in more than 60% of the instances with deployed airbag. In rear-end collisions, the airbag is rarely deployed (less than 5% of

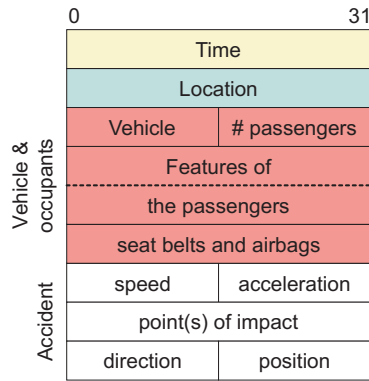


Fig. 11. Warning packet format for the proposed system.

the available instances), making the body type and the speed limit the most relevant features in the classification.

C. AI application to our proposed system

The obtained results are very useful to estimate the effectiveness of the system, as well as determining the needed data to be collected from the crashed vehicles. The presented Bayesian models generate accurate enough predictions to be used in the Control Unit of our system, which will be later applied to build resource sets adapted to each accident.

In addition, the transformation phase allows us to define the required data set that vehicles should collect and send after a collision happens, for each type of accident. We used this information to define the structure of our warning message, containing a set of fields accessible via the sensors installed within the vehicle (see Figure 11). The message structure selected can be easily adapted to match the Basic Safety Message (BSM) defined in the Society of Automotive standard J2735 [21] by means of using the Abstract Syntax Notation (ASN) encoding used for the BSM. According to the previous analysis, our warning packet includes the following information:

TIME

- timestamp with the local time, to inform exactly when the accident occurred.

LOCATION

- geographical position of the vehicle, obtained through the GPS integrated system to allow the emergency services determine the location of the crashed vehicle.

VEHICLE-OCCUPANTS

- characteristics of the vehicle, especially the body type, since its importance to determine the severity of side and rear-end collisions is proven.

- number of passengers, to adequate the medical team required to attend them.
- features of the passengers: these data is important for the rescue teams to adapt the medical equipment, but it is not decisive to determine the severity of the injuries on the passengers.
- information about seat belts and airbags, this information is crucial to estimate the severity of the injured occupants, how the accident occurred and the severity of the accident.

ACCIDENT

- speed and acceleration of the vehicle just before the accident, to estimate the severity of the accident, especially in front collisions.
- point(s) of impact, i.e. it indicates which parts of the vehicle received the impact during the accident.
- direction of impact force. Used to determine the type of accident detected (front, side, or rear-end crash).
- position of the vehicle after the crash to warn the emergency team about the level of complexity of the rescue.

The prediction models used to estimate the severity of the accident were built using the full data obtained from the GES database. However, as shown in the transformation phase, only a subset of the possible attributes was really relevant for the estimation depending on the main direction of the impact. Hence, the system should focus on obtaining at least these attributes to avoid a noticeable reduction in terms of accuracy. Additionally, since the information most complicated to be collected, i.e., the personal information such as age and sex, has little relevance from the estimation perspective compared to the rest of variables that should be directly obtained from in-vehicle sensors, we consider that the system is robust enough even when handling incomplete data. In our system, the precision of the GPS system is enough to locate crashed vehicles, since this position will only be used by the emergency services to determine the area affected by the accident. Modsching et al. [22] discovered that average urban scenarios (the most adverse scenarios for GPS positioning) produce a mean error on GPS location of about 15 meters when the street presents high buildings at both sides, but the error is reduced to just 2 meters on average when there is a clearer view of the sky, since more satellites could be used to estimate the position. Additionally, using the information contained in the built-in street maps to correct the current location of the vehicle, e.g., avoiding impossible positions inside of buildings, helps to reduce the mean error to just 5 meters on average.

VI. PROTOTYPE IMPLEMENTATION AND VALIDATION

To test the correctness of the proposed design, a prototype was built using low cost off-the-shelf devices. We measured the capacity of the system for correctly detecting and notifying the accident to the Control Unit, and the generation of appropriate severity estimations from the collected data.

A. Propotype Development

The Data Acquisition Unit of the OBU was built using an ARM microcontroller, programmed to periodically collect data from in-vehicle sensors. Basically, these sensors are accelerometers and gyroscopes that indicate the severity of the impacts received by the automobile or the occurrence of a rollover that might endanger the integrity of the occupants. Communication between the microcontroller and the Processing Unit is done by sending UDP packets through an Ethernet interface. The OBU Processing Unit in our prototype is a general-purpose Asus Eee PC netbook, equipped with a solid state disk (SSD) to minimize the possibility of damage due to impact in crash tests. The vehicle position and speed are obtained using a GPS device accessible using Bluetooth.

The Control Unit prototype for the initial tests was built using common software components, allowing fast prototyping with little cost. The reception/interpretation module was implemented using the Java programming language. This module acts as a concurrent server, creating different execution threads to handle each message received, which allows exploiting multiprocessor or multicomputer architectures. Databases are managed using the MySQL relational database management system. MySQL was selected because of its scalability and easy integration with the additional components of the Control Unit. The Web server for the visualization module is Apache. To support dynamic content, we use the PHP (Hypertext Preprocessor) technology, which is easily integrated into Apache. By combining these technologies and MySQL, users can visually check the system status.

B. Prototype Validation

The prototype was validated at the Applus+ IDIADA [23] Passive Security Department facilities in Santa Oliva (Tarragona, Spain). These facilities house one of the most sophisticated crash test laboratories in the world, and constitute an official center for approval under the Euro NCAP program. Due to the cost of using real vehicles in the collision experiments, the e-NOTIFY prototype tests were performed using a platform (known as "sled") that moves on rails in order to collide against a series of metal bars that simulate the deformation suffered by a vehicle body to absorb the impact. The speed of the stroke and the configuration of the bars used in the test determine, respectively, the kind of accident detected and the segment the simulated vehicle belongs to (family car, off-road, etc.). Tested speeds are determined by European standards, and vary from 10 km/h to 64 km/h to represent different accident severities. Figure 12 shows the sled used in the tests.

The test system included an external Road Side Unit (RSU) receiving regular information from the sled (via a wireless network) of the measurements recorded by the OBU to ensure the proper behavior of the sensor reading module, along with a Control Unit in charge of receiving alert messages and applying the corresponding algorithms. The real trials were performed with two different objectives: proving that the OBU prototype was solid enough to resist a dangerous impact, and thus that it could

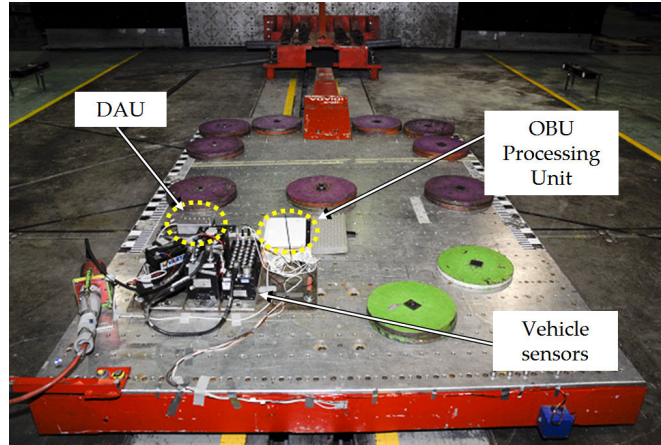


Fig. 12. Sled with the prototype installed before a crash detection test.

TABLE IV

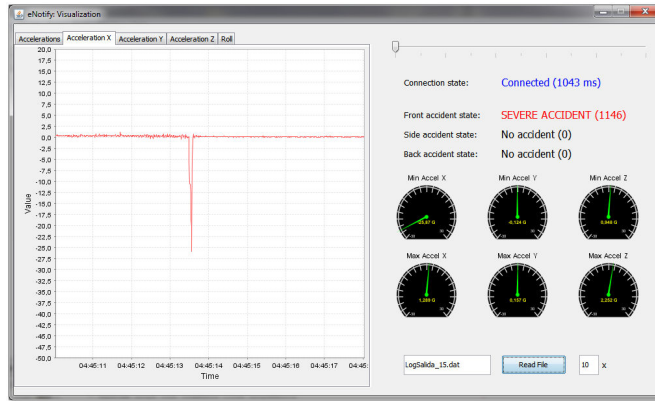
VALIDATION TESTS PERFORMED AND ACCIDENT DETECTION PARAMETERS OF THE PROPOSED SYSTEM.

Accident type	Vehicle segment	Severity	Speed	Pulse	Acceleration	Pred. Accuracy
Front accident	Large family car	Severe acc.	64 km/h	110 ms	23-28 G	80%
	Large MPV	Minor acc.	36 km/h	100 ms	15-21 G	90%
	Small family car	No acc.	15 km/h	110 ms	4-9 G	100%
Side accident	Small Off-road 4x4	Severe acc.	29 km/h	90 ms	14-21 G	80%
	Supermini	No acc.	12 km/h	90 ms	3-6 G	100%
Rear-end accident	Small MPV	Severe acc.	21 km/h	110 ms	8-12 G	73.33%
	Supermini	No acc.	10 km/h	70 ms	2-6 G	93.33%

continue working after the accident, and also ensuring the proper function of the system under a realistic crash situation, being able to determine the severity of the accident.

All the tests produced very positive results, since the OBU did not suffer from noticeable damage even in the strongest impacts (64 km/h). The durability of the equipment, based on off-the-shelf devices, was remarkable after successfully surviving unharmed to all the collision tests performed, and no error induced by the malfunction of the equipment was detected during the experiments. The detection of the accident was performed by using data from accelerometers installed on the sled. In order to avoid false detections, the system uses both the duration of the acceleration pulse, i.e., the time during which the vehicle suffered from levels of acceleration higher than 3 G, and its amplitude or peak acceleration value in the pulse. We tested several accident configurations, with different vehicle types and directions of the impact, as shown in Table IV. The values for the speed of the vehicle before the impact are defined by IDIADA, depending on the type of vehicle, direction of the impact, and severity of the collision, following assessment programs like the Euro NCAP.

After a collision is detected, the OBU builds a warning message using the data from the accelerometer and the rest of sensors available in the vehicle, following the format showed in Section V-C.



The screenshot shows a web interface for a 'Service Center' with the number '112'. It has tabs for 'Alerts', 'Filed alerts', and 'Exit'. The main content is a table titled 'Notified alerts' with the following data:

State	Geo Loc.	Date	Acc. Type	Rollover	Severity	Rescue card	File	More Info.
112	SCX	2011-05-25 13:48:58	Front (severe)	NO	!!!	🚑	📁	+
112	SCX	2011-05-24 20:34:45	Back	NO	!	🚑	📁	+
112	SCX	2011-05-11 20:34:45	Back	NO	!	🚑	📁	+
112	SCX	2011-05-11 18:33:25	Front (minor)	NO	!	🚑	📁	+
112	SCX	2011-05-11 17:58:41	Right	NO	!	🚑	📁	+
112	SCX	2011-05-06 11:15:38	Front (severe)	NO	!!!	🚑	📁	+

At the bottom of the interface, it says 'GRUPO DE REDES DE COMPUTADORES, UPV & EHU, 2011'.

Fig. 13. Images of the crash test results: Accident pulse recorded by the OBU (top), and the same accident notified and received by the CU (bottom).

Since the tests rely on a simulated vehicle, the values of some of the variables (like model year, weather, etc.) were filled with typical values. The message constructed this way was sent to the Control Unit using UMTS technology, in a similar way as if the vehicle had access to the cellular infrastructure. Finally, the Control Unit obtained the data from the warning message sent by the vehicle. Figure 13 shows a sample of the acceleration measured by the OBU in a severe front impact, and the corresponding notification received by the Control Unit containing the values collected from the vehicle on-board sensors.

The data received by the Control Unit were later inserted into the severity estimation models presented in Section V-B, based on Bayesian filters and built using historical data. We conducted several crash tests to evaluate the prediction accuracy of these prediction models. The front, side, and rear-end collisions were repeated 30 times for each type of impact while varying the accident

severity, obtaining the results shown in Table IV. The OBU was able to detect non-accident situations in almost all cases, and thus the Control Unit was not informed about those incidents. The prediction ability for rear-end collisions is certainly lower than the results achieved for the rest of accidents, since the experiments involved only one vehicle and no information about the other vehicles was simulated. However, in a real environment, the Control Unit would receive this information as well, thereby increasing the accuracy of the prediction system.

VII. RELATED WORK

Several approaches can be found in the existing literature with the objective of increasing traffic safety through the use of telecommunication technologies, and also in the field of accident severity estimation using historical data.

A. *Using Telecommunication Technologies to Improve Traffic Safety*

The U.S. Department of Transportation (DOT) developed some projects similar to ours with the goal of improving traffic safety through the use of vehicular communication, based on testing the effectiveness and safety benefits of the wireless connected vehicle technology in real-world, multimodal driving conditions [24]. Some preliminary results regarding the distance traveled by warning messages, number of relaying vehicles, and communication times in a real experiment in the streets of Los Angeles (USA) in 2011 can be found in [25]. However, these experiments only include V2V communications and the notification of dangerous situations between vehicles, whereas our system mainly concentrates on improving the decision making process that follows the occurrence of an accident. In addition, the tests were performed using the IEEE 802.11b and 802.11g standards, instead of the IEEE 802.11p standard, which was specially designed to be used in vehicular environments.

There are also some projects that make use of ECG (Electrocardiogram) sensors to monitor the drivers' condition and detect possible health problems that could endanger traffic safety.

- The authors in [26] propose a condition monitoring system to obtain physiological signals for car driver's health condition monitoring by means of ECG and PPG (Photoplethysmogram) sensors attached to the steering wheel. The signals concerning heart rate are transmitted to a server PC via a Personal Area Network (PAN) for practical tests, being analyzed to detect drowsiness and fatigue. The results obtained from this system could be used to inform nearby vehicles about a dangerous driver status, but the system does not include notification of abnormal statuses since only a PAN is considered.
- Palantei et al. [27] designed a wireless system for remotely monitoring heart pulses obtained from ECG sensors. This system allows communication of the monitoring information 50 to 250

meters far from the person, but it is not indicated how the signal is processed to classify the status of the person.

- The monitoring system proposed in [28] includes non-intrusive active electrodes installed on the seats of the vehicle. The data collected is sent through a wireless PAN and the processing of the data concerning heart rate variability in time and frequency allowed determining if the driver is tired or stressful.

We can see that these approaches are too limited by the wireless technology used, which provides a very short communication range. Moreover, they have only been used to determine the fatigue or stress level, probably due to the inviability of finding real cases to test their efficiency on the estimation of the injuries suffered after an accident. The integration of ECG sensors in modern vehicles could be an excellent opportunity to collect information about health signs after the occurrence of an accident, since our proposed architecture would allow the notification of the gathered data to the Control Unit for further processing and classification by means of intelligent algorithms.

B. Previous Approaches towards Accident Severity Estimation using Data Mining

Despite the interest that may arise from understanding the influence of various factors on road accidents, the number of works about this topic in the literature is not particularly large. In addition, most attempts to carry out a data mining process related to traffic accidents only considered data from a single city or a very small area, making results only slightly representative.

Several works are based on data obtained from the Traffic Office of Ethiopia, since this country presents one of the largest number of accidents per capita. Beshah and Hill [29] used data from 18,288 accidents around Addis Ababa as the basic data set. This study uses Naïve-Bayes, decision trees, and k-nearest neighbors (KNN) algorithms to classify the data using a cross-validation methodology, with accuracy values close to 80%. However, the authors only provided estimations for the whole accident, not for single occupants. Data from Ethiopia was also used to build regression tree models for accident classification in [30]. Only 13 out of 36 variables available in the data were used to build the classification models, but the selection process was not shown, and again only estimations about the whole accident were provided.

The area of South Korea was also selected to develop classification models based on artificial neural networks, decision trees, and logistic regression [31]. The data set involved 11,564 accidents, and the authors concluded that the different classification algorithms provide similar results in terms of accuracy, being the use of protection devices, such as the seat belt and the airbag, the most relevant factors to classify accidents. This work was extended in [32] using ensemble methods (i.e., multiple models to obtain better predictive performance than could be obtained from any of the

constituent models) combined with a prior assignation of instances through clustering, attaching a different classification model to each cluster, which produced a better class assignment.

More recently, Chong et al. [33] selected data from all over the United States obtained during the 1995-2000 period to propose a set of models based on artificial neural networks, decision trees, and Support Vector Machines (SVMs). All the classification models presented similar accuracy results, and they were highly effective at recognizing fatal injuries.

Finally, some authors have focused on the characteristics of specific road segments, instead of using the data from individual vehicles. Clustering of accident hotspots were performed by Anderson [34] in order to determine effective strategies for the reduction of high density areas of accidents. Authors studied the spatial patterns of injury related to road accidents in London (UK), and they found several hotspots with relevant significance using K-means clustering. Nayak et al. [35] used a road-based approach for modeling the crash proneness of road segments using available road and crash attributes, classifying the road depending on their "crash proneness". They also used a prior clustering of the accidents with similar features, and a consequent classification of the data by means of decision trees. However, they did not establish different severity levels on the accidents studied.

From previous works, we detected significant shortcomings when attempting to combine their results with vehicular networks, since existing works about estimating the severity of road accidents have not been used to improve the assistance to injured passengers. All the above papers used a whole variety of attributes to build the classification models, whereas only some of them could be effectively extracted from the vehicle itself (e.g., the driver's inebriation level). In addition, none of them used an adequate feature selection algorithm to select the optimal variable subset. Finally, some of the models are extensively used (decision trees), while other interesting methods received minor attention (SVMs, and Bayesian networks). To the best of our knowledge, none of these approaches have been implemented and tested in a real environment, since they only make use of historical data.

It is also noteworthy that existing proposals in the literature trying to estimate the severity of a traffic accident do not develop a complete KDD process. In fact, the only phase of the KDD process that has received widespread attention is the data mining phase, while the rest has been overlooked or summarized as much as possible. Although data mining is a very important phase, the results obtained when omitting the previous phases can lose their interest or utility.

VIII. CONCLUSIONS

The new communication technologies integrated into the automotive sector offer an opportunity for better assistance to people injured in traffic accidents, reducing the response time of emergency services, and increasing the information they have about the incident just before starting the rescue process. To this end, we designed and implemented a prototype for automatic accident notification

and assistance based on V2V and V2I communications. However, the effectiveness of this technology can be improved with the support of intelligent systems which can automate the decision making process associated with an accident. A preliminary assessment of the severity of an accident is needed to adapt resources accordingly. This estimation can be done by using historical data from previous accidents using a Knowledge Discovery in Databases process.

Most of the existing work focused on data mining in traffic accidents is based on data sets where a very limited preprocessing and transformation were performed. After a careful selection of relevant attributes, we showed that the vehicle speed is a crucial factor in front crashes, but the type of vehicle involved and the speed of the striking vehicle are more important than speed itself in side and rear-end collisions. The status of the airbag is also very useful in the estimation, since situations where it was not necessary to deploy the airbag rarely produce serious injuries to the passengers.

The studied classification algorithms do not show remarkable differences, but we demonstrate that, if we are able to classify the accidents depending on the types of impacts, we can noticeably increase the accuracy of the system, especially for front crashes where the vehicle is usually the striking one. To this end, we developed a prototype that shows how inter-vehicle communications can make accessible the information about the different vehicles involved in an accident. Moreover, the positive results achieved on the real tests indicates that the accident detection and severity estimation algorithms are robust enough to allow a mass deployment of the proposed system.

ACKNOWLEDGMENTS

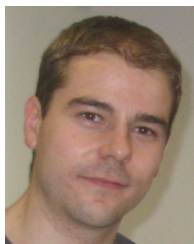
This work was partially supported by the *Ministerio de Ciencia e Innovación*, Spain, under Grant TIN2011-27543-C03-01, and by the *Diputación General de Aragón*, under Grant “subvenciones destinadas a la formación y contratación de personal investigador”.

REFERENCES

- [1] Dirección General de Tráfico (DGT), “The main statistics of road accidents. Spain,” 2010, available at http://www.dgt.es/portal/es/seguridad_vial/estadistica.
- [2] Eurostat: Statistical Office of the European Communities, “Transport statistics in the EU,” 2012, available at http://epp.eurostat.ec.europa.eu/portal/page/portal/transport/data/main_tables.
- [3] J. Miller, “Vehicle-to-vehicle-to-infrastructure (V2V2I) intelligent transportation system architecture,” in *2008 IEEE Intelligent Vehicles Symposium*, june 2008, pp. 715–720.
- [4] F. Martinez, C.-K. Toh, J.-C. Cano, C. Calafate, and P. Manzoni, “Emergency services in future intelligent transportation systems based on vehicular communication networks,” *IEEE Intelligent Transportation Systems Magazine*, vol. 2, no. 2, pp. 6–20, summer 2010.
- [5] M. Fogue, P. Garrido, F. J. Martinez, J.-C. Cano, C. Calafate, and P. Manzoni, “Prototyping an automatic notification scheme for traffic accidents in vehicular networks,” in *4th IFIP Wireless Days Conference*, October 2011.

- [6] M. Fogue, P. Garrido, F. J. Martinez, J.-C. Cano, C. T. Calafate, and P. Manzoni, "Evaluating the impact of a novel message dissemination scheme for vehicular networks using real maps," *Transportation Research Part C: Emerging Technologies*, vol. 25, pp. 61–80, December 2012, DOI: 10.1016/j.trc.2012.04.017.
- [7] B&B Electronics, "The OBD-II Home Page," 2012, available at <http://www.obdii.com>.
- [8] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "The KDD process for extracting useful knowledge from volumes of data," *Communications of the ACM*, vol. 39, pp. 27–34, November 1996.
- [9] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *SIGKDD Explorations*, vol. 11, pp. 10–18, November 2009.
- [10] European New Car Assessment Programme (Euro NCAP), "Test Procedures," 2012, available at <http://www.euroncap.com/testprocedures.aspx>.
- [11] National Highway Traffic Safety Administration (NHTSA), "National Automotive Sampling System (NASS) and General Estimates System (GES)," 2012, available at <http://www.nhtsa.gov/NASS/>.
- [12] ———, "FTP Site for the General Estimates System (GES)," 2012, available at <ftp://ftp.nhtsa.dot.gov/GES/>.
- [13] K. Yu and E. Dutkiewicz, "Geometry and Motion-Based Positioning Algorithms for Mobile Tracking in NLOS Environments," *Mobile Computing, IEEE Transactions on*, vol. 11, no. 2, pp. 254–263, feb. 2012.
- [14] M. Hall, "Correlation-based feature selection for machine learning," Ph.D. dissertation, Department of Computer Science, University of Waikato, Hamilton, New Zealand, 2008.
- [15] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence - Special issue on relevance*, vol. 97, pp. 273–324, December 1997.
- [16] J. R. Quinlan, *C4.5: programs for machine learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993.
- [17] G. F. Cooper and E. Herskovits, "A bayesian method for the induction of probabilistic networks from data," *Machine Learning*, vol. 9, pp. 309–347, October 1992.
- [18] J. C. Platt, *Fast training of support vector machines using sequential minimal optimization*. Cambridge, MA, USA: MIT Press, 1999, pp. 185–208.
- [19] T. Fawcett, "ROC Graphs: Notes and Practical Considerations for Researchers," HP Labs, Tech. Rep., 2004.
- [20] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognition*, vol. 30, pp. 1145–1159, 1997.
- [21] SAE International, "DSRC Implementation Guide – A guide to users of SAE J2735 message sets over DSRC," 2010, available at <http://www.sae.org/standardsdev/dsrc/DSRCImplementationGuide.pdf>.
- [22] Modsching, M., Kramer R., and Hagen K., "Field trial on GPS Accuracy in a medium size city: The influence of built-up," in *3rd Workshop on Positioning, Navigation and Communication 2006 (WPNC'06)*, Hannover, Germany, March 2006, pp. 209–218.
- [23] IDIADA: Instituto de Investigación Aplicada del Automóvil, "Applus+ IDIADA Web Site," 2012, available at <http://www.idiada.es>.
- [24] U.S. Department of Transportation - Research and Innovative Technology Administration (RITA), "Safety Pilot Program Overview," 2012, available at http://www.its.dot.gov/safety_pilot/.
- [25] A. Amoroso, G. Marfia, M. Rocchetti, and G. Pau, "To live and drive in I.a.: Measurements from a real intervehicular accident alert test," in *2012 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, april 2012, pp. 328–332.
- [26] H.-S. Shin, S.-J. Jung, J.-J. Kim, and W.-Y. Chung, "Real time car driver's condition monitoring system," in *2010 IEEE Sensors*, November 2010, pp. 951–954.
- [27] E. Palantei, M. Baharuddin, A. Andani, K. Nien, D. Utami, A. Febriani, U. Umar, and M. Agus, "A 2.5 ghz wireless

- ecg system for remotely monitoring heart pulses,” in *2012 IEEE Antennas and Propagation Society International Symposium (APSURSI)*, July 2012, pp. 1–2.
- [28] S. Jung, H. Shin, J. Yoo, and W. Chung, “Highly sensitive driver condition monitoring system using nonintrusive active electrodes,” in *2012 IEEE International Conference on Consumer Electronics (ICCE)*, January 2012, pp. 305–306.
- [29] T. Beshah and S. Hill, “Mining Road Traffic Accident Data to Improve Safety: Role of Road-Related Factors on Accident Severity in Ethiopia,” in *Proceedings of AAAI Artificial Intelligence for Development (AI-D’10)*, Stanford, CA, USA, March 2010.
- [30] T. B. Tesema, A. Abraham, and C. Grosan, “Rule Mining and Classification of Road Accidents Using Adaptive Regression Trees,” *International Journal of Simulation Systems, Science & Technology*, vol. 6, no. 10-11, pp. 80–94, 2005.
- [31] S. Sohn and H. Shin, “Pattern recognition for road traffic accident severity in Korea,” *Ergonomics*, vol. 44, no. 1, pp. 107–117, January 2001.
- [32] S. Y. Sohn and S. H. Lee, “Data fusion, ensemble and clustering to improve the classification accuracy for the severity of road traffic accidents in Korea,” *Safety Science*, vol. 41, no. 1, pp. 1–14, 2003.
- [33] M. Chong, A. Abraham, and M. Paprzycki, “Traffic accident analysis using machine learning paradigms,” *Informatica*, vol. 29, pp. 89–98, 2005.
- [34] T. K. Anderson, “Kernel density estimation and K-means clustering to profile road accident hotspots,” *Accident Analysis & Prevention*, vol. 41, no. 3, pp. 359–364, 2009.
- [35] R. Nayak, D. Emerson, J. Weligamage, and N. Piyatrapoomi, “Road crash proneness prediction using data mining,” in *Proceedings of the 14th International Conference on Extending Database Technology*, ser. EDBT/ICDT ’11. New York, NY, USA: ACM, 2011, pp. 521–526.



Manuel Fogue is an assistant researcher in the University of Zaragoza. He earned B.Sc. and M.Sc. Computer Science degrees from both the University of Zaragoza in 2007 and the University Jaume I of Castellon in 2009, respectively. In both cases, he graduated with honors. He received his Ph.D. degree in Computer Engineering from the University of Zaragoza in 2012. His research interests include VANET simulation, intelligent transportation systems, traffic safety, vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications.



Piedad Garrido is an assistant professor in the Department of Computers and Systems Engineering at the University of Zaragoza in Spain. She graduated in Computer Science and Documentation at the Technical University of Valencia in 1997 and 1999, respectively. She received her Ph.D. degree in Documentation from the University Carlos III of Madrid in 2008. Her current research interests include intelligent transportation systems, virtual agents, ontologies, traffic safety, vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications. She is member of the IEEE.



Francisco J. Martinez is an associate professor in the Department of Computers and Systems Engineering at the University of Zaragoza in Spain. He graduated in Computer Science and Documentation at the Technical University of Valencia in 1996 and 1999, respectively. He received his Ph.D. degree in Computer Engineering from the Technical University of Valencia in 2010, and received an Extraordinary Doctorate Award. His current research interests include VANET simulation, intelligent transportation systems, traffic safety, vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications. He is member of the IEEE.



Juan-Carlos Cano is a full professor and the Head of the Department of Computer Engineering at the Technical University of Valencia (UPV) in Spain. He earned an M.Sc. and a Ph.D. in computer science from the UPV in 1994 and 2002, respectively. From 1995 to 1997 he worked as a programming analyst at IBM's manufacturing division in Valencia. His current research interests include power aware routing protocols and quality of service for mobile ad hoc networks and pervasive computing.



Carlos T. Calafate is an associate professor in the Department of Computer Engineering at the Technical University of Valencia (UPV) in Spain. He graduated with honors in Electrical and Computer Engineering at the University of Oporto (Portugal) in 2001. He received his Ph.D. degree in Computer Engineering from the Technical University of Valencia in 2006, where he has worked as an assistant professor since 2005. He is a member of the Computer Networks research group (GRC). His research interests include mobile and pervasive computing, security and QoS on wireless networks, as well as video coding and streaming.



Pietro Manzoni is a full professor in the Department of Computer Engineering at the Technical University of Valencia (UPV) in Spain. He received the MS degree in computer science from the "Università degli Studi" of Milan, Italy, in 1989, and the Ph.D. degree in computer science from the Polytechnic University of Milan, Italy, in 1995. His research activity is related to wireless networks protocol design, modeling, and implementation. He is member of the IEEE.