A Predictive Framework of Speed Camera Locations for Road Safety

Asmae Rhanizar[1] & Zineb El Akkaoui[1]

[1] SEEDS team, STRS Lab, INPT, Rabat, Morocco

Correspondence: Asmae Rhanizar, SEEDS Team, STRS Lab, INPT, Rabat, Morocco.

Received: April 13, 2019 Accepted: May 13, 2019 Online Published: July 30, 2019

doi:10.5539/cis.v12n3p92 URL:https://doi.org/10.5539/cis.v12n3p92

Abstract

Road traffic crashes are a public health issue due to their terrible impact on individuals, communities, and countries. Studies affirmed that vehicle speed is a major contributor to crash likelihood and severity. At the same time, they identified Automated Speed Enforcement (ASE) systems, namely speed cameras, as a highly effective measure to reduce excessive and inappropriate speed, and thus improving road safety. However, identifying optimum sites for fixed speed camera placement stays an open issue in the literature, although it is a key factor that guarantees the efficiency of such ASE systems. This paper describes a predictive framework of speed camera locations using a classification algorithm that can predict, for each section of a given road network, its pertinence as a speed camera location. First, we identify a set of features as predictors of the classification algorithm, that we have argued their goodness through correlation tests. Second, for training our algorithm, data from road controlled sections, corresponding to existing speed cameras, is exploited. Each section class reflects the contribution level of the ASE system (good, neutral, or bad) to road safety. Third, as a proof-of-concept, the framework has been implemented and deployed on the Moroccan road network. The results showed that Random Forest classifier is the best performing model attaining an accuracy of 95% and a precision of 88%. Further, a tool was developed to visualize updated classification results on a Moroccan road network map to support authorities in their decision making process.

Keywords: automated speed enforcement (ASE) system, classification algorithm, road safety, speed camera

1. Introduction

According to the World Health Organization (2018), 1.35 million road traffic deaths occur every year producing a terrible impact on individuals, communities, and countries. Moreover, road traffic injury is the 8^{th} leading cause of death for all age groups especially among children and young adults aged 5 to 29 years.

To address this concern, various initiatives have emerged over the years aiming at assisting countries to increase their road safety initiatives and raise public awareness about road traffic crashes menace. An important recent initiative has been the adoption of specific targets related to road safety as part of the United Nations (UN) 2030 agenda for sustainable development. One of the new targets is to halve the global number of deaths and injuries from road traffic crashes by 2020. To promote its implementation, the UN adopted a global registry of voluntary commitments and multi-stakeholder partnerships facilitating global engagement of all stakeholders. Many international organizations such as the World Bank Group, Federation International Automobile Foundation and the International Road Assessment Program (iRAP) are partners of UN to promote road safety. In particular, the iRAP is an umbrella organization for a set of essential sub-projects, such as EuroRAP, usRAP and AusRAP. These are not-for-profit organizations dedicated to saving lives through safer roads. iRAP produces a wide range of resources and technical specifications to support road safety practitioners.

In Morocco, although the total number of registered vehicles is low compared to developed countries, the rate of road traffic deaths is still high. Indeed, 80680 crashes were recorded on Moroccan roads with 3785 deaths and 9791 people seriously injured in 2016 (Ministry of Equipment, Transport, Logistics and Water [METLE], 2016). More than two-thirds of crashes occur in built up areas, but the most severe crashes occur on rural roads where two-thirds of the fatalities are recorded (International Transport Forum [ITF], 2017). Thus, improving road safety is a top priority for the Moroccan government. In this respect, Morocco is a partner of the iRAP program to develop measures to reduce the road death toll and serious injuries in Morocco. A memorandum of understanding has been signed between the Moroccan Ministry of Equipment, Transport, Logistics and Water and the iRAP in the context of the first African road safety forum 2018, held in Morocco. Its main objective is to establish a locally-led iRAP Moroccan program.

On the other hand, according to the World Health Organization, an increase in average speed is a direct cause of both the likelihood and the severity of crashes. For example, every 1% increase in mean speed produces a 4% increase in the fatal crash risk and a 3% increase in the serious crash risk. Management of speeds of drivers of motorized vehicles has a high safety potential. Consequently, many countries have introduced an integrated approach of speed management to limit the harmful effects of excessive and inappropriate speeds in the transport system. Many measures are adopted to promote lower speeds, including automated enforcement of speed limits, road infrastructure and environment, and appropriate signing and marking (European Conference of Ministers of Transport [ECMT] , 2006).

More recently, Automated Speed Enforcement (ASE) systems are more commonly deployed along roads to help reduce driver vehicle speeds using different devices, especially fixed, mobile, and average speed cameras. Speed cameras are used to detect speeding violations by the photographic filming or videotaping of vehicles as they pass by them at a speed higher than the threshold limit. The vehicle registration number is recorded, and violation evidence is then processed and reviewed in an office environment to enable an infringement notice to be issued to the offender. ASE systems are an important element in speed management and can be a very effective countermeasure to prevent speeding-related crashes. In this context, international research based on many ASE programs clearly argues that speed cameras help change driver behavior and have a positive road safety impact (Hess, 2004; De Pauw, Daniels, Brijs, Hermans & Wets, 2014; Pilkington & Kinra, 2005; KANG, 2002).

A critical element for ASE programs is the definition of target sites where to implement speed cameras, especially fixed ones considering that they are permanently installed on predefined locations. The choice should reflect the potential of improvements due to the speed camera implementation in terms of: i.) reduction of road traffic collisions and related casualties; and ii.) reduction of average speed. Existing works formalize this problem as an operation research problem (De Leur & Milner, 2011; Boscoe-Wallace, 2017; Ko, Geedipally & Walden, 2013). However, in the best of our knowledge, few works dive into finding the best speed camera sites based on data-driven approach. The latter allows addressing site selection challenge by taking advantage of historical data. The data-driven approach consists of exploring empirical correlations and extracting patterns from past data enabling better decision making.

The objective of this paper is to address this gap by defining a predictive framework of speed camera locations. Based on a machine learning model for site selection using a classification algorithm. The framework can designate, each section of the road network, as good, neutral, or bad speed camera site. To achieve this goal, we define a Priority function as a composition of two main variables: i.) the effect of the ASE system on speed limit violation; and ii.) the effect of the ASE system on collisions of a controlled section. Then, we apply a classification algorithm by using as input the result of this priority function. As a proof of concept, the proposed framework has been implemented and deployed to predict good speed camera sites for the expansion of the Moroccan ASE program. The realized tool allows executing the classification approach and the visualization of its results on maps.

The remainder of the paper sections is as follows: The literature review is presented in section 2. Section 3 describes the approach conducted to define the priority of a speed camera site. A case study is then presented in section 4 by applying the approach proposed on Moroccan ASE program. The conclusions are given in the final section.

2. Literature Review

A large number of works (Hess, 2004; KANG, 2002; Blais & Carnis, 2015; Li, Graham & Majumdar, 2013; De Pauw, Daniels, Brijs, Hermans & Wets, 2014) have analyzed the effects of speed cameras on road safety. Most documented of them evaluate the effectiveness of ASE programs by average vehicle speeds or collisions. They have shown that the installation of speed cameras leads to a significant decrease of incidents observed in the catchment area of the camera. In addition, they demonstrated that ASE systems could reduce average speeds by at least 5km/h at the camera location.

As regards the speed camera sites selection, a set of attempts proposed various methodologies. For instance, guidance could be provided by some governments, e.g., the Province of Alberta's Automated Enforcement Guidelines (Alberta, 2014). Also, in (Newstead, 2016), locations with the highest number of total collisions are considered as the most suitable candidates for speed camera deployment. Another study was performed to identify optimal speed camera sites across New Zealand's road network. It introduced a ranking mechanism based on social costs by assigning a score to each segment of the New Zealand road network reflecting the road risk (Scott & Harris, 2016). Other works formulate the site selection process as a resource allocation and scheduling problem. For instance, a study (Kim, Wang, El-Basyouny & Fu, 2016) introduced a systematic, data-driven framework used to operate a mobile photo radar enforcement program. The site selection is based on a prioritization process informed by speed limit violation and collision data. Another study (Boscoe-Wallace, 2017) involved a model developed using optimization techniques to help predict the optimal locations for speed cameras. The developed model predicts the number of fatal and serious accidents as well as the number of slight accidents. Moreover, other types of ASE systems have also been studied. For example, a paper (De Leur & Milner, 2011) introduces a methodology for selecting good Intersection Safety Camera sites in British Columbia. The methodology is based on factors reflecting the potential of improvement of a site as well as the cost of deployment of the intersection safety camera. The literature related to applying machine learning methodology on the site selection problem for speed cameras implementation stays unfortunately, absent.

Further, machine learning models have been hugely beneficial for other transportation system issues helping to discover statistically significant patterns in multi-sourced data. Many areas are concerned especially road safety. Some studies (Mantouka, Barmpounakis & Vlahogianni, 2018; Goh, Ubeynarayana, Wong & Guo, 2018; Osman, Hajij, Karbalaieali & Ishak, 2019) have applied machine learning models to define unsafe driving behaviors. In this context, a study (Zhu, Guo, Krishnan & Polak, 2018) uses a Convolutional Neural Network (CNN) to develop an incident detection model that takes account of spatiotemporal network and traffic inherent structure in traffic incident detection. Another study (Zhang, He, Gao & Ni, 2018) applied Deep Belief Network (DBN) and Long Short-Term Memory (LSTM) deep learning models to detect traffic accidents using social media data. An additional study proposed a Chain of Road Traffic Incident Framework to predict accidents by applying a two-stage modeling procedure (Xiong, Chen & Liang, 2017). It classifies leaving lane scene versus remaining in lane scene, and then recognizes accident versus non-accident pattern given the classified scene. In other cases, machine learning techniques are used to classify crash injury severity. In fact, a study was performed to classify the level of injury severity using the Michigan Traffic Crash Facts (MTCF) dataset, from 2016 to 2017. A hybrid approach was explored to improve the crash classification performance in (Jeong, Jang, Bowman & Masoud, 2018). A study (Mafi, AbdelRazig & Doczy, 2018) involved the monetary cost of incorrect injury severity prediction through the application of C4.5, instance-based, and random forest machine-learning models.

Furthermore, some studies have introduced machine learning or deep learning models to predict road conditions. For instance, a study (Piryonesi & El-Diraby, 2018) built a decision tree model to predict the level of Pavement Condition Index (PCI) deterioration.

3. Approach

To tackle the literature gap of a data-driven methodology that deals with the site selection issue for speed cameras, we propose in the present work a machine learning approach using a classification algorithm. Such algorithm mainly needs two elements: a concise definition of the classes, and the features that will be used as descriptors of these classes.

3.1 Problem formulation

It is noteworthy that the road network is composed of rural roads, urban roads and highways. We focus the study on rural roads and highways because they share comparable features related to their geographical characteristics. We define a controlled section as a section of road where a fixed speed camera is installed. The controlled section matches the catchment area, which is the surrounding area of the camera where the effect is remarked. Note that from this point forward, fixed speed cameras will simply referred to as speed cameras.

In the subsequent, we will depict how we have identified the class and then the features considered in the approach.

3.2 Class definition

As said, the objective of this work is to determine among all sections of the road, those able to contribute the best to road safety once setting up a speed camera. To this purpose, we elaborate a classifier that classes the contribution to road safety of each road section as good, neutral or bad. Good sites are sites where a speed camera is the most effective.

To judge the section contribution, and thus its class, we establish a new measure called "Priority" (P). This measure evaluates the impact of the speed camera implementation on the safety of the controlled section by assessing its effect on collisions and speed limit violations.

Many ASE programs usually evaluate the impact of speed cameras by the frequency of collisions or crash severity on the controlled sections (De Leur & Milner, 2011; Hess, 2004). Other programs move towards evaluating the effectiveness of speed cameras by assessing the evolution of vehicle average speeds at the camera location (De Pauw, Daniels, Brijs, Hermans, & Wets, 2014; Kang, 2002). From our perspective, a speed camera is an effective road safety measure if it contributes simultaneously to decrease road collisions as well as speed limit violations. In other terms, a priority P is defined as a combination of collision gain and speed limit violation gain. Speed camera locations will be ranked then according to their Priority (P) values, computed using the formulation described below.

3.2.1 Collision Gain

To evaluate the impact of speed cameras on collisions, we inspire from the risk mapping method defined by the euroRAP program, which allows evaluating the crash risk in each road section (The European Road Assessment Programme [EuroRAP], 2018). The risk rate chosen to evaluate the gain related to collisions G_C is called the annual crash density. It shows the actual observed number of crashes per unit length and therefore where the highest and lowest numbers of crashes on the network. It is expressed as the number of fatal and serious injury crashes per kilometer per year. Thus, the annual crash density in a controlled section is defined as follows:

$$CD_{in} = \frac{FS_{in}}{L_i} \tag{1}$$

Where: $CD_{in} = Crash\ density\ of\ the\ year\ (n)\ for\ the\ site\ (i)$;

$FS_{in} = Fatal\ and\ serious\ crashes\ of\ the\ year\ (n)\ for\ the\ site\ (i)$;

$L = length\ of\ the\ controlled\ section\ of\ the\ site\ (i)$

According to the euroRAP method, to the value of crash density risk rate can belong to one of five risk banding groups. The upper and lower limits of each band, also defined by the program euroRAP, need to be normalized by the scaling factor SF_n. It is the ratio of fatal and serious crashes to fatal crashes. This approach shows that it is possible to normalize risk mapping thresholds between countries given the differences in reporting fatalities systems.

$$SF_n = \frac{FS_n}{F_n} \tag{2}$$

Where: $SF_n = Scaling\ Factor\ of\ the\ year\ (n)$; $F_n = Fatal\ crashes\ of\ the\ year\ (n)$

Table 1. Crash density bands

Position	Risk band	Lower-upper limits	Normalized lower-upper limits
1	Low	$0\ to < 0.03$	$0\ to < 0.03 \times SF_n$
2	Low-medium	$0.03\ to < 0.05$	$0.03 \times SF_n\ to < 0.05 \times SF_n$
3	Medium	$0.05\ to < 0.08$	$0.05 \times SF_n\ to < 0.08 \times SF_n$
4	Medium-high	$0.08\ to < 0.11$	$0.08 \times SF_n\ to < 0.11 \times SF_n$
5	High	≥ 0.11	$\geq 0.11 \times SF_n$

However, since we want to involve the effect on speed camera violations in the priority P, the bands will be defined later to conveniently rank speed cameras locations. For P construction, the collision gain G_C will be hence computed as the evolution of the crash density according to a base year (b) corresponding to the year where the enforcement policy isn't yet applied.

$$G_C(i) = CD_{ib} - CD_{in} \tag{3}$$

Where: $G_C(i) = Collision\ gain\ of\ the\ site\ (i)$

3.2.2 Speed Limit Violation Gain

The second gain used in P formula is related to the evolution of speed limit violations in controlled sections. We propose to calculate the daily annual number DAN of speed limit violations of each site.

$$DAN_{in} = Average\ (number\ of\ speed\ limit\ violations\ in\ each\ day\ of\ the\ year\ (n)\ for\ the\ site\ (i))$$

The speed limit violation gain is thus computed as the evolution of the DAN in a site according to a base year (b) corresponding to the year where the enforcement policy is not yet applied.

$$G_{SLV}(i) = DAN_{ib} - DAN_{in} \tag{4}$$

Where: $G_{SLV}(i) = Speed\ limit\ violation\ Gain\ of\ the\ site\ (i)$

3.2.3 Priority

Now, we will combine both speed limit violation and collision gains. We will first normalize them through a Min-max normalization to rescale their range and be able to compose them.

$$G^*_{SLV}(i) = \frac{G_{SLV}(i) - G^{min}_{SLV}}{G^{max}_{SLV} - G^{min}_{SLV}} \tag{5}$$

$$G^*_C(i) = \frac{G_C(i) - G^{min}_C}{G^{max}_C - G^{min}_C} \tag{6}$$

Where: $G^*_{SLV}(i) = Normalized\ Speed\ limit\ violation\ Gain\ of\ the\ site\ (i)$; $G^*_C(i) = Normalized\ Collision\ Gain\ of\ the\ site\ (i)$;

$G_{SLV}^{min}, G_{C}^{min} = Minimum\ values\ of\ Speed\ limit\ violation\ and\ Collision\ gains\ respectively;$

$G_{SLV}^{max}, G_{C}^{max} = Maximum\ values\ of\ Speed\ limit\ violation\ and\ Collision\ gains\ respectively$

Second, we need to combine G_C^* and G_{SLV}^* in order to rank sites according to crossing effect of the speed camera on both collisions and speed limit violation. We assume that there is no sense to take into account the collision gain alone since it could be due to other factors different than speed cameras. Thus, we draw this combination through a quadratic expression of P defined as follows:

$$\boxed{P_i = (G_C^*(i) - \overline{G_C^*}) * (G_{SLV}^*(i) - \overline{G_{SLV}^*}) + (G_{SLV}^*(i) - \overline{G_{SLV}^*})^2} \tag{7}$$

Where: $\overline{G_C^*} = Median\ of\ normalized\ collision\ gain;\ \overline{G_{SLV}^*} = Median\ of\ normalized\ speed\ limit\ violation\ gain$

Actually, the first part of the equation $(G_C^*(i) - \overline{G_C^*}) * (G_{SLV}^*(i) - \overline{G_{SLV}^*})$ represents the interaction of speed limit violation and collisions impacts, whereas $(G_{SLV}^*(i) - \overline{G_{SLV}^*})^2$ assesses the individual impact on speed limit violation. This means that a good site is where a speed camera has a positive effect on speed limit violation individually or the interaction of speed limit violation and collisions effects.

It is worth noting that P can be calculated by assigning relative weights for collisions and speed limit violation. The coefficients typically represent the cost per unit of collisions and speed limit violation respectively, as has been discussed in a previous work (Kim, Wang, El-Basyouny & Fu, 2016).

$$P_i = \alpha_i * G_C^*(i) + \beta_i * G_{SLV}^*(i) \tag{8}$$

Where: $\alpha_i, \beta_i = relative\ weight\ for\ collisions\ and\ speed\ limit\ violation\ respectively,\ of\ the\ site\ (i)$

However, for the present work, the need is to classify the sections precisely and not provide and discuss the exact relative contributions of each effect (through relative α_i and β_i) to road safety.

Finally, we define three priority bands to rank sites according to their P values. A *class* per site is deduced on a three-point scale: Bad, Neutral and Good.

3.3 Classification Features

To characterize a candidate site for implementing a speed camera, we propose the features in Table 2. The features are defined as relevant characteristics that seem to describe thoroughly the aspects related to road safety.

Table 2. Features description

Code	Feature	Description
$AADT$	AADT	Annual average of the daily traffic in the controlled section
N_{lanes}	Number of lanes	The number of lanes of the controlled section
LM	Line marking	Either the line is continuous or broken on the controlled section. Continuous line = 1; Broken line = 0
MS	Median strip	Either the controlled section has a median strip or not, knowing that a median strip is a separation between opposing traffic lanes. Has a median strip = 1; Has not a median strip = 0
SL	Speed limit	The speed limit on the controlled section
L_{ss}	Length of the straight section	The length of the straight section of the road. It's defined by calculating the distance between two curves upstream and downstream the controlled section.
HE	Highway exit	Either the controlled section is near a highway exit or not when the site is on a highway. Near a highway exit =1; Not near a highway exit = 0
HT	Highway toll	Either the controlled section is near a highway toll or not when the site is on a highway. Near a highway toll =1; Not near a highway toll = 0
RA	Rest area	Either the controlled section is near a rest area or not when the site is on a highway. Near a rest area = 1; Not near a rest area = 0
EC	Entrance to a city	Either the controlled section is near an entrance to a city. Near an entrance to a city = 1; Not near an entrance to a city = 0
D_{ua}	Distance from an urban area	Road distance from the nearest urban area

The attributes in Table 2 were determined according to their possible association either with driver's attitudes to speeding or with road safety in the controlled section. Some attributes describe road characteristics related to speed behaviors, such as road surface marking LM, traffic $AADT$, and number of lanes N_{lanes}. Others intend to detail features in relation to the position in the road. Specifically, the proximity to special areas, e.g., rest areas RA, tolls HT or exits HE on highways, urban areas UA.

To prove the relevance of the features and their correlation with the outcome *class* variable, we evaluated their dependence using statistical correlation tests as a part of the case study presented in the next section. The Pearson's coefficient was used for quantifying linear dependence for continuous variables whereas the Chi-Square was used for evaluating categorical attributes. We note that a feature extension is envisaged in the next versions of the proposed framework to cover other aspects such as weather conditions and the age of circulating vehicles on roads.

4. Case study: The Moroccan speed camera project

Speed cameras were first introduced in Morocco by 2010 as a strategic answer to road safety and ASE program requirements. More recently, Morocco has adopted by 2017 a new national road safety strategy (2017 to 2026) with the following targets:

- Reduce the number of deaths to less than 3000 fatalities by 2020;
- Reduce the number of deaths to less than 1900 fatalities by 2026.

In this purpose, Morocco aims to earn new speed cameras and responsible authorities are led to define new pertinent camera sites. At the same time, they need to evaluate the current sites if they should remain or be replaced.

4.1 Architecture overview

As a key project stakeholder, we proposed an overall predictive framework to tackle these issues. The framework constitutes a real-world implementation of the approach described in the sections mentioned above. It is implemented according to the component architecture depicted in Figure 1.

Figure 1. Component architecture

4.1.1 Road Network

The Moroccan road network is composed of speed cameras. Speed cameras detect speed limit violations according to a threshold limit, and send the reports to the processing servers.

4.1.2 Processing Servers

They are servers hosting the following databases:

- The *speed limit violations* database is recording the results of the processing of all speed limit violations detected by speed cameras.
- The *road collisions* database is reporting all injury or fatal crashes collected at the scene of the crash by police. Data related to accidents involving only material damage is not recorded in this database.

The processing servers are controlled by the Moroccan Center of Automated Offense Processing, a unit belonging to the Moroccan Ministry of Equipment, Transport, Logistics, and Water.

4.1.3 Prediction Module

It's the module used to predict locations where to implement new speed cameras. This module is implemented as a tool offering the main components described next:

- A classification component that uses the data collected from processing servers, after applying a set of necessary transformations to score any location of the road network. Specifically, this component fragments the road network into sections of the same distance such as 2 kilometer, performs the evaluation and returns an output about the goodness of each section as a new speed camera location.
- A visualization component that visualizes the classification results on a high-quality map to make it easy to be exploited by end users.

4.2 Prediction Module

In this section, we will emphasize the last part of the architecture, the prediction module, as it represents the main implementation part of our predictive framework. As said, the prediction module makes use of historical data (speed limit violations and road collisions databases), extracted from the current speed cameras locations, to predict new ones. Figure 2 shows the four steps that were undertaken while building the prediction module.

Figure 2. Prediction module

4.2.1 Data Initialization

Since the data quality and quantity directly determine the effectiveness of any data-driven framework, an important step in the prediction module is to gather the data and prepare it.

We assume that the historical data is a representative sample of the Moroccan road network as it covers different parts and regions of the road network. At the same time, the accuracy and error measures that are computed in the next step will confirm the quality of the prediction module results.

This step performs: 1.) data consolidation to fusion the two databases as mentioned earlier; 2.) data cleansing by replacing blanks and treating aberrant values; and finally 3.) the aggregation of the obtained data. Actually, data needs first to be merged conveniently by mapping attributes from both databases and other external data sources such as satellite maps, i.e., data consolidation. Data cleansing is also necessary if the data involved in the processing is incomplete or unavailable, the instance is either replaced or removed from the resulting dataset to prevent predicting errors. For instance, a speed camera can be non-operational for a period or collision data could not be available for a section of road. Data aggregation consists of a set of computations, such as computing the DAN and CD measures, defined in the Approach section. Table 3 provides a snippet of the resulting data. It is worth noting that the chosen base year 2015 belongs to a test year where no speed violation policy was applied in the studied speed camera sites although already settled and deployed. Thus, for the base year, the speed cameras have no impact on the driver's behaviors.

Table 3. The output of data initialization (snippet)

Speed camera id	DAN_b	DAN_a	CD_b	CD_a
1	1605	284	0.75	1
69	139	100	0.75	0
14	211	101	0.25	0.25
11	276	116	1.75	0.5
29	74	16	0.75	0.5
2	818	184	0.75	1

4.2.2 Data preparation This step consists of extracting features and identifying the classes (good, neutral, or bad), through the computation of the priority function for all the studied locations, necessary for the classification algorithms.

- *Computing priority*

  In order to compute priority, both G_C and G_{SLV} gains are processed using the CD_b , CD_a, DAN_b, and DAN_a values as previously detailed in the Approach section. Speed camera sites are then ranked according to their P values. Then banding thresholds are defined as a one-dimensional clustering (based on the priority measure) to identify three groups. Each of the groups corresponds to respectively good, neutral, or bad *class*.

Table 4. The output of data preparation (snippet)

$AADT$	N_{lanes}	LM	MS	SL	L_{ss}	HE	HT	RA	EC	D_{ua}	$Class$
10588	3	0	0	60	0.34	0	0	0	0	13.8	Good
10588	3	0	0	60	0.34	0	0	0	0	13.8	Good
11002.5	1.5	1	0	80	1.55	0	0	0	1	2	Bad
4785.5	1	1	0	60	1.56	0	0	0	1	1	Bad
10032.5	2	0	1	60	14	0	0	0	0	12.2	Good
60211	1.5	0	1	120	10	1	0	0	0	20	Good
15813	1	0	1	120	10	1	0	1	0	20	Bad

- *Feature extraction*

  The identified features in Table 2 were fed using the available information collected from speed limit violations database in addition to the annual traffic data and geographical maps published by the Moroccan authorities. We note that the mean of imputation corrected missing data. Table 4 shows a snippet of the resulted dataset.

Figure 3. Prediction module

We first searched for some patterns in the dataset to better describe it, that is called descriptive analysis. For example, for the line marking LM and number of lanes N_{lanes} features, we plot the contingency tabulation since they are categorical variables as shown in Figure 3. At first sight, we can notice that the proportion of bad and neutral sites is more significant for sites with a broken line and few lanes.

Moreover, to ensure that we only include variables that are good predictors of the target class, we should confirm that these variables explain an acceptable amount of variance of the outcome. Evaluation of features' correlation is operated in order to test relationships between different attributes and the target *class*. The Chi-Square was used for testing dependence between the *class* and categorical features and determining whether they are relevant to the predicted outcome. For continuous features, the correlation between the *class* and each of them was assessed through the computation of the Pearson coefficient. These tests show all the features are effectively well correlated to the target outcome.

4.2.3 Classification

As said, the site selection problem was treated in this work as a supervised classification problem. But, since they are several classification algorithms in the literature, we need to identify the best-suited algorithm to our purpose. For this reason, we executed a set of commonly known algorithms, on the obtained dataset, to evaluate and choose the one offering the best performance.

Actually, by using the Python Scikit-Learn library, several classification methods were trained, tested, and compared, including Support Vector Machine, K Nearest Neighbor, K-means, Random Forest, Gaussian Naive Bayes, and Linear Discrimination Analysis algorithms.

- *Training & Testing*

  The output dataset from the data preparation step is used to train classification models. We separated the training and test subsets of the dataset. The training set was used to fit and tune the model whereas the test set was used to evaluate the model. The prepared dataset was actually split into training and test datasets according to an 80-20% ratio, as a first step. And the classifier was trained with the eleven features. In a second step, we processed cross-validation to avoid underfitting or overfitting issues and to improve the prediction accuracy of the test data.

Table 5. Accuracy measures

Accuracy %	Using traditional training set	Using cross-validation set
Support Vector Machine	60	58
Gaussian Naive Bayes	73	81
K-nearest Neighbors	40	48
K-means	27	47
Random Forest	80	95
Linear Discriminant Analysis	33	43

To be able to select the best-suited algorithm, it is important to evaluate the performance of each model. Table 5 describes accuracy percentages using traditional training set 80-20% and using cross validation set.

Table 6. Precision, recall and F1-score measures

Measure %	Precision	Recall	F1-score
Support Vector Machine	50	60	55
Gaussian Naive Bayes	76	73	75
K-nearest Neighbors	49	40	44
K-means	26	27	24
Random Forest	88	87	87
Linear Discriminant Analysis	48	33	36

Knowing that the accuracy could be biased due to a large number of True Negatives (TN), we dive into assessing more metrics to evaluate the models better. The confusion matrix was thus used to find the correctness and accuracy of each model. It represents the ratio of the number of correct predictions to the total number of input samples. An ultimate objective during the prediction is to minimize the False Positives to minimize bad or neutral sites predicted as good sites for implementing speed cameras. Table 6 describes precision, recall and F1-score percentages by model. To our case, precision is a priority metric knowing that the costs of False Positives are high.

Figure 4. ROC curve of Random Forest classifier

On the other hand, in order to evaluate the separability of models and visualize the performance of models examined, we have plotted the ROC curve. Figure 4 shows the ROC curve of the Random Forest classifier that presents the best AUC.

Based on the above, the Random Forest algorithm was the most accurate model with an excellent generalization performance. We choose it then for building the target classifier.

- *Prediction*

  In order to predict new speed camera sites, the Morrocan road network was divided into partitions of 2 kilometer knowing that the length of partitions corresponds to the length of the catchment area of speed cameras assumed to be 2 kilometer (Hess, 2004; Kang, 2002). A new dataset was prepared to contain partitions of the Moroccan road network. Features were currently manually extracted for road partitions. The model constructed ranked the new sites according to their contribution to road safety from a speed limit enforcement perspective.

4.2.4 Representation

In order to ease the interpretation of classification results, we propose a tool that enables the visualization of good, neutral and bad speed camera sites on a map which source code is hosted in Github (Note 1). The recommended tool is able to process either the whole road network or a specific region. An output map is produced allowing to easily understand the distribution of good sites and easing decision-making for the Moroccan authorities.

Figure 5. Map representation of the classification results

Figure 5 shows the projection of the classification results on a map. On the left, a snippet of classified sections of the Moroccan Rabat-Kenitra region is depicted. A color scale was used to differentiate between bad, neutral and good sections. Good partitions are presented with a bright red (Dangerous sections) whereas bad and neutral sections are in green and blue respectively. The figure on the right shows the zoom capability of the tool that allows focusing on only two sections of the RN9 national road.

5. Conclusions and Future Work

Many international programs aim to get rid of collisions due to vehicle speed. The speed camera is a significant enforcement system deployed along roads to help reduce driver vehicle speeds. According to many research works, speed cameras improve driver behavior and have a positive road safety impact.

Predicting pertinent sites where to implement speed cameras to guarantee more road safety is not a straightforward problem. It depends on many factors depending on road conditions, such as traffic, shape, and geographical characteristics, as well as the rates of speed limit violations and collisions.

This work mainly aims at extracting the most relevant factors and using these factors to propose a well-founded formulation of the prediction problem. As a result, we established a twofold formulation as a combination of:

- A quadratic Priority function - of speed limit violation and collision gains - to determine the usefulness of speed camera enforcement in a point of the road;
- A classification problem using this priority level as the target to be predicted, and road characteristics as predictors. In the following problem, we have employed algorithms such as Support Vector Machine, K Nearest Neighbor, K-means, Random Forest, Gaussian Naive Bayes and Linear Discrimination Analysis. Best results were provided by Random Forest;

As proof of work, the established prediction solution has been deployed on the Moroccan road network and has proven a high-quality precision. A visualization tool has been developed, as well, in order to make the solution more beneficial for Moroccan authorities.

As a continuation of this work, we plan to fully automatize the road characteristics extraction by using deep learning on satellite maps, instead of the current manual extraction. Another interesting extension consists of diversifying the road characteristics taking into account, for instance, meteorological and behavioral aspects of drivers in the region.

Acknowledgments

This work has been conducted in cooperation with the Ministry of Equipment, Transport, Logistics and Water. Authors are grateful to the Directorate of Road Transport and Road Safety for their support and granting access to useful information and databases to accomplish this research work.

References

Alberta, J., & Solicitor, G. (2014). *Automated Traffic Enforcement Technology Guidelines, Province of Alberta*. Retrieved from https://open.alberta.ca/dataset/e6334b81-feea-4435-87ec-4a0cf85382e8/resource/b43f0d67-851c-4c89-83bb-7d2f0e0e904f/download/2014-automated-traffic-enforcement-technology-guidelines-sept-2014.pdf

Blais, E., & Carnis, L. (2015). Improving the safety effect of speed camera programs through innovations: Evidence from the French experience. *Journal of Safety Research*, *55*, 135-145. https://doi.org/10.1016/j.jsr.2015.08.007

Boscoe-Wallace, A. (2017). *Optimisation of speed camera locations using genetic algorithm and pattern search* (PhD thesis, Loughborough University, Leicestershire, UK). Retrieved from https://dspace.lboro.ac.uk/dspace-jspui/handle/2134/25179

De Leur, P., & Milner, M. (2011). Site Selection Process and Methodology for Deployment of Intersection Safety Cameras in British Columbia, Canada. *Transportation Research Record*, *2265*(1), 129-136. https://doi.org/10.3141/2265-14

De Pauw, E., Daniels, S., Brijs, T., Hermans, E., & Wets, G. (2014). Behavioural effects of fixed speed cameras on motorways: Overall improved speed compliance or kangaroo jumps? *Accident Analysis & Prevention*, *73*, 132-140. https://doi.org/10.1016/j.aap.2014.08.019

European Conference of Ministers of Transport (ECMT) (2006). *Speed Management*, The Organisation for Economic Co-operation and Development (OECD) Publishing, Paris. https://doi.org/10.1787/9789282103784-en

Goh, Y. M., Ubeynarayana, C. U., Wong, K. L. X., & Guo, B. H. W. (2018). Factors influencing unsafe behaviors: A supervised learning approach. *Accident Analysis & Prevention*, *118*, 77-85. https://doi.org/10.1016/j.aap.2018.06.002

Hess, S. (2004). Analysis of the Effects of Speed Limit Enforcement Cameras: Differentiation by Road Type and Catchment Area. *Transportation Research Record*. https://doi.org/10.3141/1865-05

International Transport Forum (ITF) (2017). *Road Safety Annual Report 2017*, OECD Publishing, Paris. https://doi.org/10.1787/irtad-2017-en

Jeong, H., Jang, Y., Bowman, P. J., & Masoud, N. (2018). Classification of motor vehicle crash injury severity: A hybrid approach for imbalanced data. *Accident Analysis & Prevention*, *120*, 250-261. https://doi.org/10.1016/j.aap.2018.08.025

KANG, J. (2002). Changes of speed and safety by automated speed enforcement systems. *IATSS Research*, *26*, 38-44. https://doi.org/10.1016/S0386-1112(14)60041-8

Kim, A. M., Wang, X., El-Basyouny, K., & Fu, Q. (2016). Operating a mobile photo radar enforcement program: A framework for site selection, resource allocation, scheduling, and evaluation, Case Studies on Transport Policy. *Case Studies on Transport Policy*, *4*, 218-229. https://doi.org/10.1016/j.cstp.2016.05.001

Ko, M., Geedipally R. S., & Walden, D. T. (2013). Effectiveness and Site Selection Criteria for Red Light Camera Systems. *Transportation Research Record: Journal of the Transportation Research Board*, *2327*, 53-60. https://doi.org/10.3141/2327-07

Li, H., Graham, D. J., & Majumdar, A. (2013). The impacts of speed cameras on road accidents: An application of propensity score matching methods. *Accident Analysis & Prevention*, *60*, 148-157. https://doi.org/10.1016/j.aap.2013.08.003.

Mafi, S., AbdelRazig, Y., & Doczy, R. (2018). Machine Learning Methods to Analyze Injury Severity of Drivers from Different Age and Gender Groups. *Transportation Research Record: Journal of the Transportation Research Board*, *2672*, 171-183. https://doi.org/10.1177/0361198118794292

Mantouka, E. G., Barmpounakis, E. N., & Vlahogianni, E. I. (2018). Mobile Sensing and Machine Learning for Identifying Driving Safety Profiles. *Transportation Research Board 97th Annual Meeting*. Washington DC, United States.

Ministry of Equipment, Transport, Logistics and Water (METLE). (2016). *Recueil des statistiques des accidents corporels de la circulation routire 2016*. Retrieved from http://www.equipement.gov.ma/AR/Infrastructures-routieres/Reseau-Routier-du-Maroc/Documents/Recueil%20des%20accidents%20de%202015%20VF.pdf

Newstead, S. V. (2016). Outcomes of a workshop to determine criteria for placement of fixed speed cameras and red light speed cameras under the automated traffic enforcement program in Western Australia. Perth, Western Australia. Retrieved from https://www.rsc.wa.gov.au/RSC/media/Documents/Resources/Publications/criteria-for-placement-of-fixed-speed-cameras.pdf

Osman, O. A., Hajij, M., Karbalaieali, S., & Ishak, S. (2019). A hierarchical machine learning classification approach for secondary task identification from observed driving behavior data. *Accident Analysis & Prevention*, *123*, 274-281. https://doi.org/10.1016/j.aap.2018.12.005

Pilkington P., & Kinra S. (2005). Effectiveness of speed cameras in preventing road traffic collisions and related casualties: systematic review. *BMJ*, *330*, 330-331. https://doi.org/10.1136/bmj.38324.646574.AE

Piryonesi, M., & El-Diraby, T. (2018). Using Data Analytics for Cost-Effective Prediction of Road Conditions: Case of The Pavement Condition Index. *US Department of Transportation, Federal Highway Administration Research and Technology*. Retrieved from https://www.fhwa.dot.gov/publications/research/infrastructure/pavements/ltpp/18065/18065.pdf

Scott, N., & Harris, D. (2016). Identifying optimal sites for static speed cameras in New Zealand: a geospatial approach. *Proceedings of the 2016 Australasian Road Safety Conference*. Canberra, Australia

The European Road Assessment Programme (EuroRAP). (2018). *RAP road risk mapping manual: technical specification*. Retrieved from https://www.irap.org/2013/08/risk-mapping-technical-specification/

World Health organization (WHO). (2018). *Global status report on road safety 2018*. Retrieved from https://www.who.int/violence_injury_prevention/road_safety_status/2018/en

Xiong, X., Chen, L., & Liang, J. (2017). A New Framework of Vehicle Collision Prediction by Combining SVM and HMM. *IEEE Transactions on Intelligent Transportation Systems*, *19*, 699-710. https://doi.org/10.1109/TITS.2017.2699191

Zhang, Z., He, Q., Gao, J., & Ni, M. (2018). A deep learning approach for detecting traffic accidents from social media data. *Transportation Research Part C: Emerging Technologies*, *86*, 580-596. https://doi.org/10.1016/j.trc.2017.11.027

Zhu, L., Guo, F., Krishnan, R., & Polak, J. W. (2018). The Use of Convolutional Neural Networks for Traffic Incident Detection at a Network Level. *Transportation Research Board 97th Annual Meeting*. Washington DC, United States

Notes

Note 1. https://github.com/INPT-SEEDS/PF

Copyrights