

Accident black spots identification based on association rule mining

Abdelilah Mbarek¹, Mouna Jiber², Ali Yahyaouy¹, Abdelouahed Sabri¹

¹LISAC Laboratory, Faculty of Sciences Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, Fez, Morocco

²Ministry of Equipment and Water, Fez, Morocco

Article Info

Article history:

Received Mar 3, 2023

Revised Aug 27, 2023

Accepted Dec 13, 2023

Keywords:

Black spot

Data mining

Road accident

Road safety

Weighted severity index

ABSTRACT

This paper presents an analytical approach to identifying the important characteristics of accident black spots on Moroccan rural roads. An association rule mining method is applied to extract road spatial characteristics associated with fatal accidents. The weighted severity index was calculated for each section, which was then used to determine the severity levels of black spots. The apriori algorithm is applied to find the correlation between road characteristics and the severity levels of black spots. Then, a general rule selection method is proposed to identify the rules strongly associated with each severity level. The results show that the proposed approach is effective in identifying the most important factors contributing to accidents. Furthermore, it shows that the combination of several road characteristics, such as road width, road surface, and bridge presence, may contribute to fatal accidents. The general rule selection found that wet, bad surfaces, and narrow shoulders were significantly associated with accidents on rural roads. The findings of the present study can help develop effective strategies to reduce road accidents and thus improve road safety in the country.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Abdelilah Mbarek

LISAC Laboratory, Faculty of Sciences Dhar El Mahraz, Sidi Mohamed Ben Abdellah University
Fez, Morocco

Email: abdelilah.mbarek@hotmail.com

1. INTRODUCTION

In response to the resurgence of road traffic accidents, Morocco has been a forerunner in developing a national road safety strategy with the primary goal of reducing the number of deaths and serious injuries in a sustainable and continuous manner. Until the creation of the national road safety agency NARSA in 2018, the country's road safety sector was managed by the Ministry of Equipment and Water (MEW). Given the importance of road infrastructure and its role in the road safety sector, the road infrastructure budget accounts for 15% of total public investment. In 2018, more than 24 billion dirhams were invested in transportation infrastructure, with roads and highways receiving 52% of the funds. According to the same year's data, the national road network consists of 1,800 kilometers of highways and 57,334 kilometers of roads, 44,180 kilometers of which are paved, counting 77% of the entire network. The pavement condition indicator (PCI) indicates that 62.7% of the road network is in acceptable or good condition. Despite significant investments in road safety by the country, the number of fatal accidents continues to rise, resulting in significant human and economic losses for victims, their families, and the country as a whole.

In 2017, Morocco recorded 3,274 fatal accidents causing 3,726 deaths and more than 10 thousand serious injuries. According to MEW statistics, fatal accidents are recorded in good road and weather conditions, i.e., well-paved roads, dry surfaces between 6 and 10.5 meters wide, and in normal weather.

Countermeasures and prevention strategies can be developed by studying the factors that directly affect fatal accidents [1]. In Morocco, as elsewhere in the world, road crashes are killing more and more people worldwide and they are mainly due to a combination of three factors: human, road, and vehicle. Traffic accidents are a major challenge in the management of road traffic, frequently occurring in congested areas where drivers try to maintain a high speed to avoid traffic jams [2]. Apart from factors related to driver behavior or vehicle condition, the spatial characteristics of roads are considered a key element to be studied in order to prevent incidents or traffic jams. Taking into account the road safety aspect in the design and improvement of infrastructures is becoming more and more of a priority for the actors of this sector. Thus, many studies have shown that the spatial composition of certain road sections can have a close link with the type and number of accidents that occur there. Therefore, the combination of several characteristics may contribute to the occurrence of certain fatal accidents on a recurring basis. The sections concerned are often called road black spots. Their analysis is essential to improve road safety. However, it happens that some road sections with similar characteristics are not declared fatal because they have not recorded a minimum number of fatalities and/or injuries. To avoid this confusion and especially to anticipate possible black spots, it is necessary to analyze the main characteristics common to these roads, considered dangerous. This would help stakeholders in their decisions regarding the construction of road infrastructure as well as in the improvement of existing roads.

The infrastructure and road environment has been studied through various approaches to extract new knowledge in specific cases. The study by Mbarek *et al.* [3] used analysis of variance (ANOVA) to examine the relationship between road infrastructure characteristics and road mortality rates, aiming to identify the key factors that contribute to fatal road accidents. The results of the study indicated that road infrastructure characteristics have a significant impact on road mortality rates. The study by Babić *et al.* [4] examined previous studies on the effects of road markings on driver behavior and road safety. They found that wide markings reduce accidents and have a positive effect on road safety, especially when they are well-maintained and have higher retro-reflectivity. Another study showed that road infrastructure improves road safety and road user behavior [5]. It has concluded that the benefits of helmet laws are reinforced by the improvement of road infrastructure. Helmet laws are associated with an increase in helmet use among cyclists and a decrease in head injuries and deaths among motorcyclists and cyclists. The study conducted by Ambros *et al.* [6] analyzed the impact of shoulder widths on crash frequency. Several studies focused on single-vehicle accidents, such as Liu and Xia [7] who investigated fatal accidents involving a single vehicle in metropolitan areas. The study carried out by Bisht and Tiwari [8] evaluated the impact of roadside characteristics on fatal crashes on rural roads. Some studies analyzed driver behavior in specific situations. Such as the study by Van Treese *et al.* [9] who studied the impact of the presence of trees along roads on driver behavior and psychology. The study by Ben-Bassat and Shinar [10] analyzed driver perception and behavior in different shoulder, guardrail, and roadway geometry situations. Another study by Calvi *et al.* [11] analyzed driving behavior on 2+1 roads. They showed that the type of median has a significant effect on the distance between the vehicle and the median, with a greater distance when the median is a cable barrier. However, the type of median does not affect the speed in the passing lane.

Siregar *et al.* [12] evaluated the impact of the driving environment on pedestrian accidents. The results of the study showed that driving environment risk factors significantly increase the probability of fatal injury in a pedestrian accident. The study conducted by Gherghina *et al.* [13] examines the link between road transport infrastructure and gross domestic product per capita (GDP per capita). The results show that infrastructure positively influences GDP per capita. Thus, results on the impact of investments in transport infrastructure on economic growth show a positive impact.

Machine learning is one of the methods used to analyze the relationship between road infrastructure and road accidents. The study by Siregar *et al.* [14] analyzed the effect of road geometry on rural fatal accidents. Three machine learning methods were evaluated and compared for predicting fatal accidents. They showed that road features are important in explaining variations in road accidents. Another study showed an indirect relationship between road geometry and accidents. The study carried out by Siregar *et al.* [15] found that road geometric characteristics had indirect effects on the number of fatal accidents but had a direct effect on speed and speed deviations. In fact, an increase in road geometry resulted in a decrease in speed. The study by Ashraf *et al.* [16] analyzed wrong-way driving accidents using machine learning methods and an interpretable machine learning technique. Association rules are common and well-studied methods in data mining, intending to discover statistically significant associations between two or more variables recorded in massive databases. In the field of road safety, most of these techniques focus on identifying factors that affect the severity of an accident.

Recent studies have employed association rule mining methods to identify spatially related infrastructure and accident environment factors. These studies have analyzed multi-factor fatal accidents and demonstrated that infrastructure and environment have an impact on accidents [17], [18]. The study by Xu *et*

al. [19] investigated vehicle, driver, geometry, and road environment risk factors in fatal crashes with more than 10 deaths using association rules and descriptive statistics. The study carried out by Das *et al.* [20] examines driving behavior in foggy environments, specifically lane-keeping skills. They revealed that infrastructure influences motorist behavior negatively. Although the infrastructure of road intersections was not reviewed, Shahin *et al.* [21] investigated intersection incidents and concluded that building speed humps can minimize the number of major accidents. Because of the different methods and algorithms suggested in the literature, knowledge extraction utilizing association rule techniques is getting increasingly difficult and time-consuming. To overcome these difficulties, numerous approaches have been proposed for analyzing road accident data. The study conducted by Zeinab *et al.* [18] utilized *electre tri* and *promethee* for the selection of interesting and relevant rules from the vast number of rules generated. The study by Gu *et al.* [17] employed the lift increase tests to eliminate redundant rules. The trial-and-error approach involves continuously iterating the combination of support and confidence probabilities and determining a set of relatively high thresholds to ensure the credibility of the association rules. The study conducted by Xu *et al.* [19] chose the support, confidence, and lift thresholds based on previous studies. They utilized relatively high threshold values intending to increase the accuracy and strength of the generated association rules. The study by Das *et al.* [20] employed the lift increase threshold to select rules with antecedents that contribute significantly to the strength of the rule. Then, rules with more antecedents were selected over simpler rules based on the lift increase criteria.

Following our examination of the state of the art in accident data analysis, we discovered that earlier work concentrated on combining several components, such as driver, pedestrian, and weather-related aspects. Yet, infrastructure and road environment variables have received little attention. This encourages us to perform a more in-depth analysis. This includes additional features such as left and right shoulders, as well as their width and level relative to the road. In this context, we propose an analytical approach to draw out important information about the characteristics of black spots to prevent road disasters. The approach is based on association rule mining (ARM), a well-known data mining technique that has proven itself in several fields, including road safety. These techniques are well at identifying the correlation between various characteristics of road accidents, which can be difficult to analyze using traditional regression methods. These methods use accident data to generate rules about the common characteristics of black spots. Moreover, ARM methods generate rules that are easy to interpret and understand, which is not the case for AI methods where the output is not easily interpretable. This is particularly important in the context of fatal road accidents, where stakeholders need to understand the factors that contribute to accidents and deaths. Interpretability is essential for ensuring that the insights generated from the analysis are actionable. For this purpose, the *apriori* algorithm extracts the association rules between the spatial characteristics of roads and the severity level of black spots. Indeed, the aim is to find the common features of a particular type of section. In this study, four types of levels, previously determined from the results of the weighted severity index method, are considered (level 1, level 2 level 3, and safe). The primary objective of considering different types of sections was to: i) identify the common pattern of the most and least fatal sections, ii) efficiently save and manage the budget contributed to the road infrastructure, and iii) minimize ineffective interventions and inappropriate road infrastructure improvements.

2. METHOD

2.1. Study design

This study was conducted on the basis of accident data from Morocco for the period (2016-2017). We conducted this study based on fatal accidents in rural areas. Including, vehicle-vehicle, single-vehicle, and vehicle-pedestrian accidents. The components of this study consist of a health survey of the victims and a road section examination that is conducted by qualified personnel. The health survey is conducted through personal interviews, and the examination is conducted by direct measurement. Data is then processed following the steps outlined in Figure 1.

2.2. Data preprocessing

The ministry of equipment and water collects and processes statistical forms on road traffic accidents through the general directorate of roads and land transport. For each injury accident, information describing it is collected by the royal gendarmerie and services under the general directorate of national security intervened in the accident location. This information is collected in a form called road traffic accident reports in accordance with international definitions of victims. The dataset collected from the accident analysis bulletin includes concise summaries of all personal injury accidents that occurred over two years. This information includes the location and circumstances of the accident, as well as the drivers, vehicles, and victims involved. Preliminary data processing is essential to generating powerful predictive models. It consists of data preparation, anomaly detection, and data cleaning. It mainly involves the

elimination of noise, missing values, and irrelevant attributes. This study considers spatial factors, including road information such as number, length, and type. Environmental characteristics and the state of the infrastructure in the surroundings of the accidents are also considered. Table 1 lists the relevant spatial factors and their descriptions.

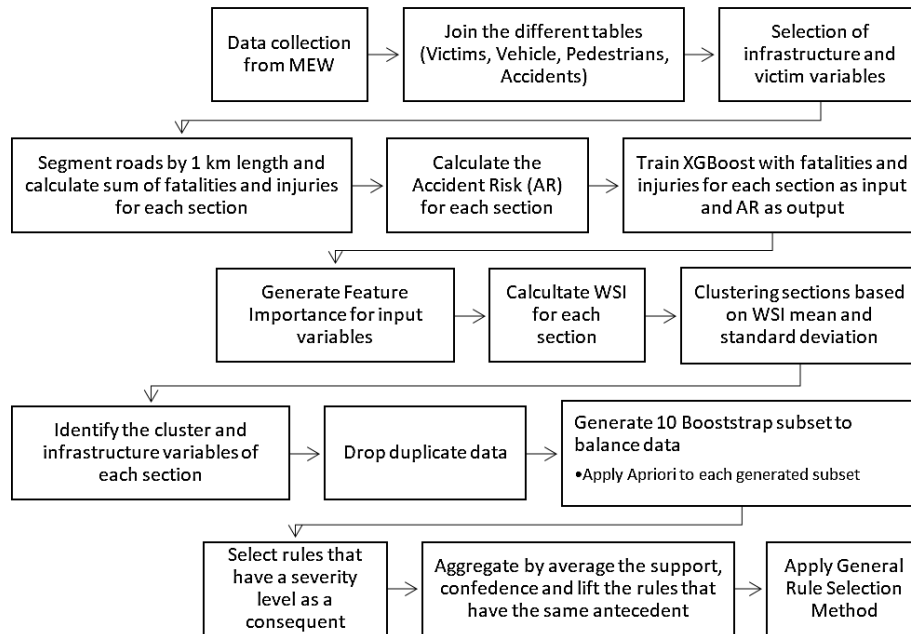


Figure 1. Summary of data preparation and analysis process

Table 1. Road spatial factors with their corresponding values

Factors	Description
Road section	Kilometer point
Road pavement condition	Good Bad Very bad
Surface condition	Unpaved road Dry Wet Snowy or icy
Road pavement width (RP)	Humid road, salted or sanded [0,2] [2,5] [5,8] [8,12] [12,90]
Road pavement type	Full road Separated by a median
Right shoulder width (RS) (resp. Left shoulder width (LS))	[0; 0.5] [0.5; 1] [1; 1.5] [1.5; 4] [4; 10]
Right shoulder position (resp. Left shoulder position)	At road pavement level Not at road pavement level
Road topography	Bridge Narrowing Obstacle
Length profile of the road	Flat Sloping Summit of a hill
Road plan layout	Straight Left-hand curve Right-hand curve Double bend

2.3. Risk of black spots

To assess the black spot risk of accidents, the weighted severity index (WSI) risk score was used, as revised in 2021 [22]. The WSI was intended to be a straightforward representation of the risk of black spots. In other words, as the WSI of a section increases, so does the risk of black spots. It is based on the mean and standard deviation of the WSI. The concept is simple: if the WSI is below a specific threshold, the risk is relatively low, and the location can therefore be considered safe. However, if the WSI exceeds a specific threshold, the risk is highly increased, and the location can be classified as excessively hazardous. WSI was calculated based on immediate fatalities, fatalities beyond 7 days, minor injuries, and serious injuries. Four categories of risk of black spots were determined, as shown in Table 2.

Table 2. Description of road black spot severity levels

WSI	Risk of black spots
$\text{mean} + 2 \text{ S.D.} \leq \text{WSI}$	High
$\text{mean} + 1.5 \text{ S.D.} \leq \text{WSI} < \text{mean} + 2 \text{ S.D.}$	Medium
$\text{mean} + \text{S.D.} \leq \text{WSI} < \text{mean} + 1.5 \text{ S.D.}$	Low
$\text{WSI} < \text{mean} + \text{S.D.}$	Safe

The preprocessed dataset contains 11702 sections of a one-kilometer length. They consist of 506 sections of high severity (level 1), 692 of medium severity (level 2), 760 of low severity (level 3), and 9744 sections that represent safe sections. Then, we proceeded to eliminate the entries that had the same characteristics and level of severity. Data redundancy may reduce the accuracy and reliability of results and increase execution times. This helps ensure that the model will perform better and produce more relevant and meaningful results. After that, we proceed to decode categorical variables. The decoding consists in assigning each numeric value, its real string value. Continuous variables, i.e., pavement and shoulder width, are then converted into categorical variables as shown in Table 1. As a result, a total of 1306 records are left, with 71 high severity sections, 79 medium severity sections, 81 low severity sections, and 1075 safe sections. Therefore, we faced an unbalanced problem that affects the generation of association rules. Unbalanced data can have a negative impact on ARM. It can lead to biased results, as the model will be more likely to identify rules that are more frequently occurring in the larger class. This can result in rules that are not representative of the true relationships in the data. Furthermore, the model may be more likely to overfit the majority class, as it has more examples to learn from.

2.4. Association rule mining

The association rule mining was first introduced by [23]. It consists of two main steps. The first step is to find all frequent itemsets that satisfy the minimum support. The second step is the generation of association rules from the frequent itemsets found in the previous step. An association rule is of the form $A \rightarrow B$, where $A \cap B = \emptyset$, A represents the antecedent, and B the consequent. The apriori algorithm is the most well-known ARM algorithm that belongs to unsupervised machine learning, which quantifies the correlation of the most frequent occurrences of items with high support, confidence, lift, and conviction, showing that each occurs frequently within a given group [24]. The Apriori algorithm was designed to operate in this study on a dataset containing spatial features of black spots.

Let D accidents dataset of n road section. Let C a set of road characteristics $C = \{c_1, c_2, \dots, c_m\}$ and $S = \{\text{Level1}, \text{Level2}, \text{Level3}, \text{Safe}\}$ a set of road section severity levels, let A a set of characteristics where $A \subset C$ and B an item of black spots severity levels where $B \subset S$ and $A \cap B = \emptyset$. There were 3 metrics used for evaluating discovered rules:

The support for the rule $(A \rightarrow B)$ is the probability that the two behaviors are more likely to occur together, where A is the antecedent and B is the consequent. It is calculated as (1):

$$\text{Support}(A \rightarrow B) = P(A \cap B) = \frac{\text{No of sections having A and B}}{\text{Total number of sections}}, \text{ range: } [0,1] \quad (1)$$

The confidence for the rule $(A \rightarrow B)$ is the conditional probability of having severity level B given that a road section has an A item. The confidence is 1 for a rule $(A \rightarrow B)$ if the consequent and antecedent always occur together. It is expressed by (2):

$$\text{Confidence}(A \rightarrow B) = \frac{P(A \cap B)}{\text{Support}(A)}, \text{ range: } [0,1] \quad (2)$$

The lift of the rule ($A \rightarrow B$) is the confidence of the rule divided by the support of the consequent B. A higher lift value between two variables means a higher correlation between them. Note, $\text{lift} > 1$ means that antecedent and consequent are more likely to appear together, while $\text{lift} < 1$, means that they are more likely to appear separately. Finally, $\text{lift} = 1$ means that there is no association between the antecedent and the consequent. It is calculated by (3):

$$\text{Lift}(A \rightarrow B) = \frac{P(A \cap B)}{\text{Support}(A) \times \text{Support}(B)} = \frac{\text{Confidence}(A \rightarrow B)}{\text{Support}(B)}, \text{ range: } [0, \infty] \quad (3)$$

2.5. Bootstrap resampling method

The data obtained from the previous step exhibited an imbalanced distribution, with a lower-level class significantly outnumbering the other classes. This is expected given that sections deemed as “safe” are more abundant than those considered risky. To address this issue, the bootstrap resampling method was employed to balance the data. The method was utilized to generate subsets from the risk class of black spots, ensuring that all classes were of equal size to the smallest class. The bootstrap subset was extracted through a random resampling process with replacement. This process was repeated multiple times to generate subsets of equal size.

In fact, 10 subsets were generated. These subsets were utilized for association rule generation. To ensure equal representation of all severity classes, the number of records for each class was set equal to that of the smallest class (level 1, with 71 instances). The Apriori algorithm was then applied to each subset using the following parameters: minimum support of 0.01, a minimum lift of 1.1, minimum confidence of 0.75, and a minimum length of 2 for association rules consequent. The combined results of these ten subsets produced 373402 rules, many of which were duplicated. To mitigate this issue, the duplicates were aggregated using the mean, resulting in 151671 unique rules. To determine the most influential rules, the rules with a minimum lift of 3 and confidence of 0.8 were selected. At the end of this step, the number of rules generated was 468.

2.6. General rules selection

The selection of association rules is a crucial step in the process of knowledge discovery in data mining. The goal of this method is to uncover relationships between items within a transaction dataset. To determine the most general association rules, the evaluation criteria of support, confidence, and lift are commonly utilized. Association rules with the highest values for these measures are deemed the most reliable and therefore chosen for further knowledge discovery.

Let D be the set of association rules. The association rules to be deleted from D assume that for two rules: $R_1: A_1 \rightarrow C$ and $R_2: A_2 \rightarrow C$ such that $A_2 \subset A_1$ and $\text{Lift } R_1 \leq \text{Lift } R_2$, thus delete R_1 . The set S defined in this way contains the association rules of set D that have no lower lift overrule in set D. Thus, S contains the most general association rules of set D. The set of general rules is defined by (4):

$$S = \{A \rightarrow C / \nexists A_1 \in D \text{ such that } A_1 \subset A \text{ and } \text{Lift } A \leq \text{Lift } A_1\} \quad (4)$$

3. RESULTS AND DISCUSSION

The results of the general rules selection are presented in Tables 3 to 6. These rules will allow us to understand the relationship between the road characteristics and the severity level of road sections. The generated association rules have revealed various factors associated with black spots of accidents. The association rules for the level 1 section (Table 3) have revealed that dry roads, with a width of 8 to 12 meters, featuring a bridge, and having a right shoulder with a width of 0.5 to 1 meter are common to all the generated rules. Thus, sections with such characteristics are dangerous in terms of mortality. The support value of 0.0211 indicates that in 2.11% of the recorded accidents, these characteristics and the high level of severity occur together. The confidence value of 1 means that in 100% of the accident cases when the section characteristics are present, high-severity accidents will also occur. This value indicates that there is a 100% chance of having fatal accidents if the topography is a bridge, the road condition is good, the road is dry, and flat, with a width between 8 and 12 meters, and the right shoulder with a width between 0.5 and 1 meter. The lift value of 4 indicates that the above characteristics and the high level of severity are strongly correlated with each other, i.e., there is more than 4 times the chance that a level 1 classified section and these characteristics will occur together than what is expected.

The rules generated by the general rule selection method encompass risk factors such as bridge, right shoulder width ranging from 0.5 to 1 meter, roadway width ranging from 8 to 12 meters, and dry roadway surface. A similar result from [25] demonstrated that roadway width, right shoulder width, and the presence of bridges have a significant effect on the severity level of black spots. These results are also

consistent with the strategy of Morocco's road directorate for black spot identification. The strategy identified sections that are of good quality, with an undivided roadway width that further varies between 7 and 16 meters, and shoulders that are level with the roadway and are flat and straight. The identified sections also feature a dry surface at 95% and bridges at 48%. Ashraf *et al.* [16] also support our findings. They showed that roadway length and lane width were positively associated with wrong-way driving crash segments, which negatively impacts road safety and increases the number of fatal accidents. In our study, we have found that the absence of median separation in good road conditions (good pavement, dry, and straight) is associated with black spots of level 1. This discovery has also been reported by Calvi *et al.* [11] which states that the type of median separation has a negative effect on the driver's secure behavior. They showed that the type of median has a significant effect on the distance between the vehicle and the median, namely a large distance when the median separation is a cable barrier. We have found that straight sections were associated with black spots. This result is consistent with that of [14] who demonstrated that straight segments were found to be more likely to increase the number of fatal accidents than curves. Also, the results of [19] showed that the straight road segment is a fatal accident risk factor.

Table 3. General rules generated for sections with level 1 severity

Antecedents (level 1)	Supp	Conf	Lift
Flat, Dry, RS_0.5_1, RP_8_12, Good, Bridge	0.02113	1	4
Straight, RS_On_level , Dry, RS_0.5_1, LS_On_level , RP_8_12, Bridge	0.02113	1	4
Straight, RS_On_level , Dry, RP_8_12, RS_0.5_1, Bridge, Full_Road	0.02113	1	4
Straight, Dry, RS_0.5_1, RP_8_12, Good, Bridge	0.02113	1	4
Straight, Dry, RS_0.5_1, RP_8_12, Bridge	0.02113	0.9	3.62
Straight, Dry, LS_On_level , RP_8_12, RS_0.5_1, Bridge, Full_Road	0.02113	1	4
Straight, Dry, Bridge, RP_8_12, RS_On_level , Flat, RS_0.5_1	0.02113	1	4
Straight, Dry, Bridge, RP_8_12, LS_On_level , Flat, RS_0.5_1	0.02113	1	4
Straight, Dry, Bridge, RP_8_12, Full_Road, Flat, RS_0.5_1	0.02113	1	4
Dry, Bridge, RP_8_12, LS_On_level , RS_On_level , Flat, RS_0.5_1	0.02113	1	4
Dry, Bridge, RP_8_12, Full_Road, RS_On_level , Flat, RS_0.5_1	0.02113	1	4
Dry, Bridge, RP_8_12, Full_Road, LS_On_level , Flat, RS_0.5_1	0.02113	1	4

Table 4. General rules generated for sections with level 2 severity

Antecedents (level 2)	Supp	Conf	Lift
RS_On_level , Dry, RS_0.5_1, Bridge, RP_5_8, Good, Full_Road	0.0158	0.833	3.33
RS_On_level , Dry, LS_On_level , RS_0.5_1, RP_5_8, Good, Bridge, Full_Road	0.0176	1	4
LS_1.5_4, RS_On_level , Wet, Bad	0.0340	0.936	3.745
LS_1.5_4, LS_On_level , Wet, Bad	0.0340	0.936	3.745
LS_1.5_4, RP_5_8, Wet, Bad	0.0305	0.929	3.718
Left-hand_Curve, Dry, LS_On_level , RS_0.5_1, Good, Bridge	0.0176	1	4
Dry, Bridge, RS_1.5_4, Separated, LS_On_level , Flat, RP_5_8	0.0158	0.828	3.314
Dry, Bridge, RS_1.5_4, Separated, Flat, RP_5_8	0.0202	0.802	3.208

Table 5. General rules generated for sections with level 3 severity

Antecedents (level 3)	Support	Confidence	Lift
Full_Road, Very_Bad	0.01525	0.80793	3.231

Table 6. General rules generated for safe sections

Antecedents (Safe)	Supp	Conf	Lift
RS_0.5_1, RS_On_level , Flat, Left-hand_Curve, Narrowing	0.01760	1	4
RS_0.5_1, Flat, Dry, LS_On_level , Narrowing, RP_5_8, Good	0.01760	1	4
RS_On_level, Humid road, salted or sanded	0.01760	1	4
LS_0.5_1, RS_0.5_1, RS_On_level , Flat, Narrowing, RP_5_8	0.01760	1	4
LS_0.5_1, RS_0.5_1, RS_On_level , Narrowing, RP_5_8, Good	0.01760	1	4
LS_0.5_1, RS_0.5_1, RS_On_level , Dry, Left-hand_Curve, Narrowing	0.02112	1	4
LS_0.5_1, RS_Not_on_level , Straight, LS_Not_on_level	0.01760	1	4
LS_0.5_1, RP_2_5, RS_0.5_1, RS_On_level , Dry, Flat	0.01584	0.87	3.5
LS_0.5_1, RP_2_5, RS_0.5_1, LS_On_level , Flat	0.01643	0.87	3.47

The level 2 (Table 4) sections mainly concern wet and poor road conditions with a width of 5 to 8 meters, which is smaller than that of level 1. This result is consistent with that of [7], [18] who found that among the attributes that influence the severity of road accidents in black spots, there are wet roads, particularly in the case of single-vehicle accidents. This is also in line with the findings of Das *et al.* [20],

who indicate that infrastructure characteristics, including wet road conditions, have a detrimental effect on driver behavior. On the other hand, the results indicate that wet and poorly maintained roadways, with a left shoulder width between 1.5 and 4 meters and either a left shoulder or a right shoulder at road level or a roadway width between 5 and 8 meters, are more hazardous. Moreover, flat, dry, separated sections, with a roadway width between 5 and 8 meters and containing a bridge are more hazardous (sup=0.0202; conf=0.8020; and lift=3.21). Also, if such sections are further characterized by a left shoulder at road level, then the probability that they will be classified as level 2 of severity increases (sup=0.0158; conf=0.8286; and lift=3.31).

Surprisingly, bridges were found to be a risk factor in the most deadly roadway sections (level 1 and level 2). One possible explanation for this is that road bridges crossing rivers often experience a certain period of fog, which reduces driver visibility and increases the braking distance of the vehicle [26]. Another possible explanation is that bridges are sometimes subjected to wind, which results in poor lane control and vehicle loss of control. To address this issue, stakeholders recommend installing wind barriers, which are commonly adopted to improve traffic safety on bridges [27]. On the other hand, bridges were also found among the most deadly sections in urban areas [28].

The results of the general rule selection approach yielded a single rule that associates road characteristics with the level 3 sections as shown in Table 5. Such sections are characterized by an undivided roadway and very poor road surface conditions. The measurement criteria indicate that the probability of a section characterized by such characteristics being classified as level 3 is 0.8. Also, there are more than 3 times the chance that a section with these characteristics will be classified as level 3 than what is expected (sup=0.0152; conf=0.8079; and lift=3.23). These results align with the findings of [12], who showed that road conditions and status are statistically significant. Particularly, poor road condition increases the relative risk of fatal accidents.

In the safe sections (Table 6), the generated rules indicate that these sections are particularly characterized by characteristics that were not identified in the other severity levels, but not limited to, road narrowing, left and right curves, slopes, and humid, sandy or salty roads, with a road width between 2 and 8 meters, and not less than 2 meters or greater than 8 meters, a left shoulder width between 0.5 and 1 meter, the position of the right (resp. left) shoulder relative to the pavement.

The results identified only one rule containing the risk factor related to curves among the three levels (level 1, level 2, and level 3). In contrast, many rules have associated the curves and the safe class. This result is in line with Siregar *et al.* [15] study which showed that the increase of curves in a road leads to a decrease in traffic speed and, consequently, the mortality and accident rate. This can be explained by the fact that curves in rural areas may contribute to road safety by positively impacting driver behavior (reducing speed, improving anticipation of potential hazards) and increasing visibility for drivers (creating clear views and eliminating blind spots). However, this result is contradictory to those of [12], which suggested that curves present an increased risk of fatal and severe injuries in a pedestrian-involved accident.

On the other hand, results show that road slope is an insignificant factor in accident black spots; i.e. slopes and secure areas are associated. This corresponds to the conclusions of [15], which state that a larger number of slopes on a road reduces traffic speed, leading to lower and converging speeds, and therefore lower mortality and accident rates. In regards to the impact of shoulder width on road safety, our findings concur with the results of Ambros *et al.* [6] which demonstrated that shoulder width does have an effect on road safety. They found that wider shoulders were associated with accident severity. Also with the findings of Ben-Bassat and Shinar [10] who studied the impact of shoulder width on road safety. They found that shoulders do have a significant effect on actual speed, lane position, and perceived safe speed when guardrails are present. In addition, shoulder width regulations have a positive impact on road safety in terms of speed control and lane position maintenance. Also, the results of Singh Bisht and Tiwari's [8] study align with our study's results. They reported that wide shoulders presented a higher risk of fatal accidents than shoulders between 1.0 and 1.5 m wide.

An unexpected finding was that road narrowing was associated with safe sections in both good and poor road surface conditions. This can be attributed to the fact that road narrowing has a positive impact on driver behavior by encouraging caution and awareness and limiting driving speed. Furthermore, road narrowings can calm traffic and improve visibility at intersections. This finding is consistent with that of Sołowczuk and Kacprzak [29], who showed that the installation of road narrowings along with other safety features can enhance road safety at intersections. However, this result is at odds with the findings of the study by Sharma *et al.* [30], which indicated that narrow bicycle lanes were a hindrance to satisfaction with the use of such lanes, having a negative impact on the safety of cyclists.

The scatter plots displayed in Figure 2 shows the distribution of general association rules using support and confidence value. It allows us to quickly visualize the association rules that have high support and confidence value. The points at the top right of the diagram represent association rules that have high

support and confidence values, indicating a strong association between the antecedent and the consequent. The points at the bottom left of the diagram represent association rules that have low support and confidence values, indicating a weak association. The size of the points changes according to the lift value, a large size indicates a large lift value, and vice versa. Figure 2(a) shows the sections characterized by: on the one side, the road narrowing, the right shoulder between 0.5 and 1 meter, and the road width between 5 and 8 meters. On the other side, in addition to the previous characteristics, a dry surface and a left shoulder between 0.5 and 1 meter, are strongly associated with the safe sections. Figure 2(b) shows that the sections characterized by: wet surface, poor pavement surface, left shoulder between 1.5 and 4 meters, and either with shoulders at the same level of the pavement or a pavement width between 5 and 8 meters, these sections are strongly associated with the black spots of the second level of severity.

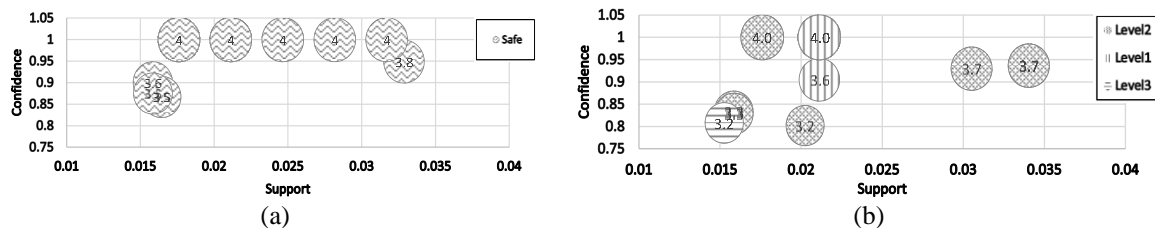


Figure 2. Scatter plot visualization of the generated rules for four black spot severity levels; (a) safe and (b) level 1, level 2, and level 3

4. CONCLUSION

This study proposed an analytical approach using ARM to identify the correlation between road spatial features and black spots in rural roads of Morocco. The method involved calculating the accident risk indicator for each 1-kilometer road section, training an XGBoost model, extracting the weights associated with each input, calculating the WSI, and grouping the sections into four levels based on WSI values. The data was then transformed into association rules, and the most general rules were selected through the proposed approach. The final step was to identify the predominant frequent features for each severity level of black spots.

The results showed that several road spatial features such as road width, road surface, and presence of bridge can contribute to the occurrence of fatal accidents in a specific type of road section. The results also demonstrate that various road features have a significant impact on the severity of black spots. The contribution of this study lies in the identification of the road features that contribute to black spots of accidents. These results can be used by road authorities to design and construct safer roads, prioritize road safety interventions, and develop effective strategies. Further studies are recommended to validate these results and examine other factors that may contribute to road accidents, such as driver behavior and vehicle characteristics. The results of this study can also be used as a basis for developing a predictive model for road accidents, which would help to identify high-risk road sections and reduce the number of accidents.

ACKNOWLEDGEMENTS

We are grateful to The General Directorate of Roads, which belongs to the Ministry of Equipment and Water, Morocco, for their valuable insights and guidance throughout this research project.




REFERENCES

- [1] "Ministère de l'Équipement et de l'Eau," 2021, Available online: <http://www.equipement.gov.ma/AR/Pages/Accueil.aspx>, date accessed: Nov. 20, 2021.
- [2] M. Jiber, A. Mbarek, A. Yahyaouy, M. A. Sabri, and J. Boumhidi, "Road traffic prediction model using extreme learning machine: The case study of tangier, morocco," *Information (Switzerland)*, vol. 11, no. 12, pp. 1–15, Nov. 2020, doi: 10.3390/info11120542.
- [3] A. Mbarek, M. Jiber, A. Yahyaouy, and A. Sabri, "Road traffic mortality in Morocco: Analysis of statistical data," in *2020 International Conference on Intelligent Systems and Computer Vision, ISCV 2020*, IEEE, Jun. 2020, pp. 1–7, doi: 10.1109/ISCV49265.2020.9204325.
- [4] D. Babić, M. Fiočić, D. Babić, and T. Gates, "Road Markings and Their Impact on Driver Behaviour and Road Safety: A Systematic Review of Current Findings," *Journal of Advanced Transportation*, vol. 2020, pp. 1–19, Aug. 2020, doi: 10.1155/2020/7843743.
- [5] B. Du *et al.*, "Deep irregular convolutional residual LSTM for urban traffic passenger flows prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 972–985, Mar. 2020, doi: 10.1109/TITS.2019.2900481.




- [6] J. Ambros, Z. Krivánková, R. Zůvala, K. Bucsuházy, and J. Frič, "Analysis of Safety Impact of Paved Shoulder Width on Czech Secondary Roads," *Archives of Transport*, vol. 60, no. 4, pp. 125–136, Dec. 2021, doi: 10.5604/01.3001.0015.6148.
- [7] X. Liu and J. C. Xia, "Locally analysing the risk factors for fatal single vehicle crashes hot spots in Western Australia," *International Journal of Crashworthiness*, vol. 20, no. 6, pp. 524–534, Nov. 2015, doi: 10.1080/13588265.2015.1055649.
- [8] L. S. Bisht and G. Tiwari, "Assessing the Effect of Paved Shoulder Width on Road Traffic Crash Risk on a Six-Lane Rural Highway in India," *Journal of Transportation Engineering, Part A: Systems*, vol. 149, no. 7, Jul. 2023, doi: 10.1061/jtepbs.teeng-7296.
- [9] J. W. Van Treese, A. K. Koeser, G. E. Fitzpatrick, M. T. Olexa, and E. J. Allen, "A review of the impact of roadway vegetation on drivers' health and well-being and the risks associated with single-vehicle crashes," *Arboricultural Journal*, vol. 39, no. 3, pp. 179–193, Jul. 2017, doi: 10.1080/03071375.2017.1374591.
- [10] T. Ben-Bassat and D. Shinar, "Effect of shoulder width, guardrail and roadway geometry on driver perception and behavior," *Accident Analysis and Prevention*, vol. 43, no. 6, pp. 2142–2152, Nov. 2011, doi: 10.1016/j.aap.2011.06.004.
- [11] A. Calvi, S. D. Cafiso, C. D'Agostino, M. Kieć, and G. Petrucci, "A driving simulator study to evaluate the effects of different types of median separation on driving behavior on 2 + 1 roads," *Accident Analysis and Prevention*, vol. 180, p. 106922, Feb. 2023, doi: 10.1016/j.aap.2022.106922.
- [12] M. L. Siregar, R. J. Sumabrata, A. Kusuma, O. B. Samosir, and S. N. Rudrokasworo, "Analyzing driving environment factors in pedestrian crashes injury levels in Jakarta and the surrounding cities," *Journal of Applied Engineering Science*, vol. 17, no. 4, pp. 482–489, 2019, doi: 10.5937/jaes17-22121.
- [13] Ş. C. Gherghina, M. Onofrei, G. Vintilă, and D. Ş. Armeanu, "Empirical evidence from EU-28 countries on resilient transport infrastructure systems and sustainable economic growth," *Sustainability (Switzerland)*, vol. 10, no. 8, p. 2900, Aug. 2018, doi: 10.3390/su10082900.
- [14] M. L. Siregar, T. Tjahjono, and N. Yusuf, "Predicting the Segment-Based Effects of Heterogeneous Traffic and Road Geometric Features on Fatal Accidents," *International Journal of Technology*, vol. 13, no. 1, pp. 92–102, Jan. 2022, doi: 10.14716/ijtech.v13i1.4450.
- [15] M. L. Siregar, T. Tjahjono, and Nahry, "Endogenous relationship of accident occurrence with speed, traffic heterogeneity and driving environment on inter-urban roads in Indonesia," *Journal of Applied Engineering Science*, vol. 18, no. 4, pp. 475–484, 2020, doi: 10.5937/jaes0-25837.
- [16] M. T. Ashraf, K. Dey, and S. Mishra, "Identification of high-risk roadway segments for wrong-way driving crash using rare event modeling and data augmentation techniques," *Accident Analysis and Prevention*, vol. 181, p. 106933, Mar. 2023, doi: 10.1016/j.aap.2022.106933.
- [17] C. Gu, J. Xu, C. Gao, M. Mu, E. Guangxun, and Y. Ma, "Multivariate analysis of roadway multi-fatality crashes using association rules mining and rules graph structures: A case study in China," *PLoS ONE*, vol. 17, no. 10, p. e0276817, Oct. 2022, doi: 10.1371/journal.pone.0276817.
- [18] F. Zeinab, K. Ali, D. Bassam, and C. Pierre, "An Improved Approach to Analyze Accidents and Promote Road Safety using Association Rule Mining and Multi-Criteria Decision Analysis Methods," *Recent Advances in Computer Science and Communications*, vol. 13, no. 4, pp. 731–746, Oct. 2019, doi: 10.2174/2213275912666190807113914.
- [19] C. Xu, J. Bao, C. Wang, and P. Liu, "Association rule analysis of factors contributing to extraordinarily severe traffic crashes in China," *Journal of Safety Research*, vol. 67, pp. 65–75, Dec. 2018, doi: 10.1016/j.jsr.2018.09.013.
- [20] A. Das, M. M. Ahmed, and A. Ghasemzadeh, "Using trajectory-level SHRP2 naturalistic driving data for investigating driver lane-keeping ability in fog: An association rules mining approach," *Accident Analysis and Prevention*, vol. 129, pp. 250–262, Aug. 2019, doi: 10.1016/j.aap.2019.05.024.
- [21] M. Shahin *et al.*, "Cluster-Based Association Rule Mining for an Intersection Accident Dataset," in *2021 International Conference on Computing, Electronic and Electrical Engineering, ICE Cube 2021 - Proceedings*, IEEE, Oct. 2021, pp. 1–6, doi: 10.1109/ICECube53880.2021.9628206.
- [22] A. Mbarek, M. Jiber, A. Yahyaouy, and A. Sabri, "A new model for black spots identification using Weighted Severity Index," in *5th International Conference on Intelligent Computing in Data Sciences, ICDS 2021*, IEEE, Oct. 2021, pp. 1–7, doi: 10.1109/ICDS53782.2021.9626758.
- [23] R. Agrawal, T. Imieliński, and A. Swami, "Mining Association Rules Between Sets of Items in Large Databases," *ACM SIGMOD Record*, vol. 22, no. 2, pp. 207–216, Jun. 1993, doi: 10.1145/170036.170072.
- [24] M. Harahap, A. M. Husein, S. Aisyah, F. R. Lubis, and B. A. Wijaya, "Mining association rule based on the diseases population for recommendation of medicine need," *Journal of Physics: Conference Series*, vol. 1007, no. 1, p. 012017, Apr. 2018, doi: 10.1088/1742-6596/1007/1/012017.
- [25] A. Mbarek, M. Jiber, A. Yahyaouy, and A. Sabri, "Black spots identification on rural roads based on extreme learning machine," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 3, pp. 3149–3160, Jun. 2023, doi: 10.11591/ijece.v13i3.pp3149-3160.
- [26] Y. Dai, Y. Bian, Y. Wu, X. Zhao, and J. Zhang, "A variable speed limit model for freeway bridges considering the influence of both traffic service level and visibility in fog," *Physica A: Statistical Mechanics and its Applications*, vol. 602, p. 127590, Sep. 2022, doi: 10.1016/j.physa.2022.127590.
- [27] Y. Zou *et al.*, "Analysis of the effects of wind barrier on driving safety and comfort of vehicles on long-span bridges under crosswinds," *Structures*, vol. 42, pp. 367–385, Aug. 2022, doi: 10.1016/j.istruc.2022.05.098.
- [28] X. Lyu and Q. Wang, "Analysis of Traffic Accident Severity on Circular Expressway Based on Logistic," *Wuhan Ligong Daxue Xuebao (Jiaotong Kexue Yu Gongcheng Ban)/Journal of Wuhan University of Technology (Transportation Science and Engineering)*, vol. 46, no. 3, pp. 411–415, 2022, doi: 10.3963/j.issn.2095-3844.2022.03.006.
- [29] A. B. Solowczuk and D. Kacprzak, "Effect of Lane Narrowing Implemented on an Urban Dual Carriageway on Speed Reduction and Acoustic Environment," *Buildings*, vol. 12, no. 1, p. 31, Dec. 2022, doi: 10.3390/buildings12010031.
- [30] B. Sharma, H. K. Nam, W. Yan, and H. Y. Kim, "Barriers and enabling factors affecting satisfaction and safety perception with use of bicycle roads in Seoul, South Korea," *International Journal of Environmental Research and Public Health*, vol. 16, no. 5, p. 773, Mar. 2019, doi: 10.3390/ijerph16050773.

BIOGRAPHIES OF AUTHORS






Abdelilah Mbarek    is a Ph.D. student in the Department of Computer Science, LISAC laboratory at the University of Sidi Mohamed Ben Abdellah, Fez, Morocco. He holds a Master's degree in Imagery and Business Intelligence in 2018, from the Faculty of Sciences Dhar El Mahraz, Fez. His current research interests include machine learning, intelligent transportation systems, road safety, road traffic prediction, and traffic accident analysis. He can be contacted at email: abdelilah.mbarek@hotmail.com.






Mouna Jiber    received her Ph.D. in computer science from the University of Sidi Mohamed Ben Abdellah, Fez in 2021. She holds a Master's degree in computer science from Alakhawayn University in Ifrane, Morocco. She currently holds the position of Head of Service at the Ministry of Equipment and Water in Fez, Morocco. Her current research interests include machine learning, intelligent transportation systems, and road safety. She can be contacted at email: m.jiber@mtpnet.gov.ma.



Ali Yahyaouy    obtained his Ph.D in Computer Science from the University of Sidi Mohamed Ben Abdellah, Fez in 2010. He is currently a Professor of Computer Science at the Faculty of Sciences Dhar El Mahraz, Fez. He is also a member of the LISAC Laboratory. He coordinates the Web Intelligence and Data Science master's program. His research focuses on MAS, ITS, reinforcement learning, machine learning, and traffic control. He can be contacted at email: ali.yahyaouy@usmba.ac.ma.



Abdelouahed Sabri    works at the Department of Computer Science, Faculty of Sciences, University of Sidi Mohamed Ben Abdellah, Fez. He does research in information science, distributed computing, and databases. His current project is 'Development of a smart learning system'. He is a member of the LIAN laboratory. He received his Ph.D. degree in 2009 from the Faculty of Sciences, Fez, Morocco. He can be contacted at email: abdelouahed.sabri@usmba.ac.ma.