International Journal of Electrical and Computer Engineering (IJECE)
Vol. 13, No. 3, June 2023, pp. 3149~3160
ISSN: 2088-8708, DOI: 10.11591/ijece.v13i3.pp3149-3160

Black spots identification on rural roads based on extreme learning machine

Abdelilah Mbarek, Mouna Jiber, Ali Yahyaouy, Abdelouahed Sabri
LISAC Laboratory, Faculty of Sciences Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, Fez, Morocco

Article Info

Article history:

Received Jun 17, 2022
Revised Oct 12, 2022
Accepted Dec 2, 2022

Keywords:

Black spots
Extreme learning machine
Road accidents
Road safety
Weighted severity index
XGBoost

ABSTRACT

Accident black spots are usually defined as road locations with a high risk of fatal accidents. A thorough analysis of these areas is essential to determine the real causes of mortality due to these accidents and can thus help anticipate the necessary decisions to be made to mitigate their effects. In this context, this study aims to develop a model for the identification, classification and analysis of black spots on roads in Morocco. These areas are first identified using extreme learning machine (ELM) algorithm, and then the infrastructure factors are analyzed by ordinal regression. The XGBoost model is adopted for weighted severity index (WSI) generation, which in turn generates the severity scores to be assigned to individual road segments. The latter are then classified into four classes by using a categorization approach (high, medium, low and safe). Finally, the bagging extreme learning machine is used to classify the severity of road segments according to infrastructures and environmental factors. Simulation results show that the proposed framework accurately and efficiently identified the black spots and outperformed the reputable competing models, especially in terms of accuracy 98.6%. In conclusion, the ordinal analysis revealed that pavement width, road curve type, shoulder width and position were the significant factors contributing to accidents on rural roads.

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

As reported by the World Health Organization (WHO), road accidents continue to be one of the leading causes of death worldwide, with serious costs for both life and the economy. Every day, they cause thousands of deaths and injuries, accounting for 2% of all deaths. According to statistics, low- and middle-income countries account for the vast majority of road traffic fatalities. In Morocco, more than 3,000 people lose their lives in traffic accidents per year during the last decade. Since 2008, the rate of accidents has continuously climbed, increasing by more than 38% between 2008 and 2017. In 2016, and 2017, the number of accidents and injuries increased by 10.8% and 9.1%, respectively, while the number of fatalities remained stable at around 3,500 [1]. As a result, the Ministry of Transport of the country established a new strategy in 2017 with the objective of reducing the mortality rate to 50% by 2026. This strategy entails strengthening the entire road network, especially the most hazardous sections. It also entails investigating black spots and determining the root causes of accidents. In this context, the motivation of the present study is to be in conformity with the government's strategy by proposing an accurate model for black spot identification.

A black spot is defined as a segment of road that has had a certain number of accidents over a given time period "*t*" and distance "*d*". In Morocco, a 1 km long road section is classified as high risk if it has at least 10 accidents causing at least 10 fatalities or serious injuries. Implementing safety measures in these sections is cost-effective and limited, due to the lack of budget. To this end, the high-risk sections should be prioritized when planning safety measures, taking into account budget restrictions. Furthermore, countries take the necessary steps to reduce the number of these black spots in light of their socioeconomic cost, particularly in terms of human life loss. There are three main methods for identifying black spots: screening methods, clustering methods, and crash prediction methods [2]. A screening method is a procedure for examining a road system in order to identify and classify potential locations from most likely to least likely in order to reduce the frequency of accidents on the roadway. Locations discovered to be the most vulnerable are thoroughly investigated in order to identify road accident patterns, risk factors, and effective safety improvement. In its most basic form, a screening method is a simple measurement that yields a “yes/no” response. It is composed of many statistical methods such as the accident rate method [3], [4], the accident frequency method [5], [6], or the accident number method. The most commonly worldwide used method is the accident number based.

As for clustering, it is the task of subdividing a data set into several groups, known as clusters, so that the elements in each group have common properties. Clusters are simply groups of data points arranged so that the distance between them is as short as possible. This type of research enables both black spot detection and accident causation analyses. However, these methods largely overlook the unpredictability and discrete nature of accidents. In terms of prediction approaches, several statistical techniques such as machine learning, predictive modeling, and data mining are explored. They estimate or forecast future outcomes using past and present data [7]. There are two types of predictive models: classification models, which predict the class of an object, and regression models, which are used to predict a number. Algorithms are then used to construct these models. Data mining, predictive modeling, and statistical analysis are performed by algorithms to extract insight from data. Ensemble methods in machine learning are multi-model systems that combine multiple models into a predictive model. They are widely used since they reduce the variance of predictions and minimize bias in the predictive model. Generally speaking, an ensemble model can be classified into two types: sequential and parallel methods. In a parallel ensemble, each individual base predictor is trained independently of the others and is generated in a parallel order with no data dependencies. Unlike parallel methods, which take advantage of each base predictor's independence, sequential methods take advantage of the base predictors’ dependence. Specifically, sequential ensembles train a new base predictor during the learning process in a way that minimizes the errors caused by the base predictor trained in the previous stage. Another way to distinguish between ensemble methods is to consider the two types of basic predictors: homogeneous and heterogeneous predictors. Heterogeneous base predictors use a variety of learning methods, whereas homogeneous base predictors use several similar predictors.

Bootstrap aggregating (Bagging) is a powerful parallel ensemble machine learning method that generates multiple versions of a predictor and then combines them to create a single aggregated prediction. It was designed to improve the performance of machine learning algorithms [8] by taking advantage of the Bootstrap method, which generates the bootstrap of the training set and uses them as new training sets for the predictors. Bootstrap data means that the samples are randomly selected with replacement, which means that the same example may be re-selected multiple times in a single sample of the training dataset. It is also possible that some examples are not selected at all for some bootstrap samples [9]. The bagging method can be used for both classification and regression problems. In the regression case, it averages the versions of the predictor, while it performs a plurality vote (voting) in the case of classification. Previous research has established that the Bagging method achieves the best improvement using only 10 bootstrap replications and that more than 25 bootstrap replications are a desperate effort [8].

In the field of road traffic data analysis, neural networks have recently received a lot of interest thanks to their simple structure, speed, and generalization performance. Extreme learning machine (ELM), a type of single hidden layer feedforward neural network (SLFN), in particular, has been widely used as a strong classifier for various benchmark data [10]. The efficacy of such a method in road traffic analysis is evaluated in a recent paper [7]. In their investigation study, different variables were identified as being associated with road traffic, and their model was able to model the traffic data with superior generalization performance thanks to this technique. Numerous studies have been undertaken and are now being conducted in the field of black spot identification. The goal is to carry out research and put it into practice to minimize the number of accidents. However, the number of in-depth studies undertaken on the spatial factors of traffic accidents remains limited. Bayesian methods are widely used in this area. They consist of combining sample data with other relevant data from the sites. Ghadi and Török [11] have revealed that these methods have flaws in their implementation, particularly on high-speed roadways. One of the most often utilized number-based strategies for identifying accident black spots is the weighted severity index (WSI) or accident point weightage (APW). The procedure comprises weighing bodily incidents to give the more serious ones a higher weight. Researchers offered numerous weighting strategies using this method. However, the arguments for the weighting strategy remain unclear and unconvincing.

Early examples of research into accidents' black spots include [12], [13]. Recently, significant analysis and discussion on the subject were presented by Mbarek *et al.* [14], who analyzed data on road accidents that were recorded on Moroccan roads during 2017. The results of this work showed that the type of intersection and the location variables are the main factors of the road conditions, that contribute most to road accident fatality. Manaa *et al.* [15] used a study to evaluate different algorithms and techniques based on computer vision and artificial intelligence (AI) to monitor and detect traffic violations and outlined their advantages and disadvantages. To date, several studies have investigated road accidents, and various of them have assessed the efficacy of the empirical Bayesian technique on black spot classification. Zhang *et al.* [16] proposed a Bayesian network to construct a black spot identification model, their binary classification model shows the best accuracy compared to the ID3 decision tree, logistic regression, and support vector machine (SVM). In addition, the results of the correlation study show that the accident location type, accident type, time, and responsibility had a significant effect on accident black spots. However, they ignored spatial variables that may have a significant impact. In [17], to classify the collected segments by severity degree, the empirical Bayesian technique was used. The decision rules technique was used to determine the major attributes that lead to accidents in specific segments. The findings show that the geometric characteristics of the road segments have a significant effect on the severity of the accident. The findings also demonstrated that the lower the risk score of a road segment, the closer will be to safe and non-hazardous environmental conditions.

A large and growing body of accident black spots research has focused on identifying and evaluating the WSI. Iqbal *et al.* [18] used APW to identify black spots, which consider fatal, major, minor, and property-damage accidents. APW and WSI have the same concept, they both weight accidents by numbers. However, the choice of weights is not justified. The main conclusion of their study is that human and vehicle errors, as well as environmental factors, are major contributors to road accidents. Calsavara *et al.* [19] studied black spots on a road with several S-curves and frequent fog. The study was conducted through a Weighted Severity Index methodology. Fatalities were weighted at 13, injuries at 5, and property damage at 1. The results showed that the identification was reliable thus contributing to the safety of the identified segments. Another previous study combined WSI and XGBoost to improve black spot identification [20]. The objective of their study is to evaluate the deadliest sections of the road to get an idea of the level of safety. In their study, the WSI method was used to calculate the severity of road sections. The XGBoost method uses statistical data on the number of deadliest accidents and casualties for each section to determine the WSI weights. The results show that XGBoost reliably determines the WSI weights. Recent studies have shown the effectiveness of the feature importance tool in feature selection. Abdu-Aljabar and Awad [21] used an importance score to extract the relevant factors for lung cancer detection. The calculated scores measure the utility of each feature in the construction of the trees which are then used in feature selection. In fact, the dimensionality reduction did not affect the performance of the model. The XGBoost model is widely used in many fields, including medicine [21]–[24], agriculture [25], the service sector [26], as well as industry sector [27].

A lot of researchers have paid particular attention to Poisson-Tweedie distribution in accident analysis [28], [29]. Debrabant *et al.* [28] used historical hospital records to identify black spots. They use a Poisson-Tweedie autoregressive to model the expected number of accidents, which accommodates both zero inflation and overdispersion and covers a wide range of discrete distributions. The data used, on the other hand, lacked credible information on local risk factors like geometry. Nguyen *et al.* [30] proposed a framework based on potential savings in accident costs or safety potential (SAPO). They proceed to Poisson testing for selected spots with high accident frequency, then calculating SAPO and ranking of spots by SAPO. The Safety potential is used as a fundamental element for identifying and ranking black spots in the technique of detecting black spots based on possible accident cost savings. Many alternative approaches have relied on geographic information systems (GIS). Rvyas *et al.* [31] used GIS for analyzing and prioritizing detected black spots. Zhu *et al.* [32] used a mathematical and a machine learning model to investigate the relationship between traffic accidents and spatial characteristics of the road. The authors propose a method to identify the accident black spots by using a GIS.

In the literature, many works study the impact of weather and vehicle conditions, as well as road user behavior, on road accident black spots. However, no research on the use of road conditions and environmental factors in the classification of accident black spots has been conducted to our knowledge. In this context, this research focuses on identifying and classifying accident black spots on Moroccan rural roads through an analysis of accident data during 2016 and 2017. Two types of fatalities are considered: on-the-spot and off-the-spot, as well as two types of injuries: serious and minor. As for the accident black spot classification part, it considers information about the location of accidents, such as road infrastructure and environment. Indeed, to our knowledge, although some studies have been conducted on accident analysis in the country, no study has been found to introduce a classification of accident black spots using infrastructure and environmental data. This is the main motivation for this work. The remainder of the paper is arranged as. Section 2 describes the used methodology. Then, we present the analysis results and discussion in section 3, and finally, we conclude the paper in section 4.

2. METHOD

The main objective of this research is to assess road and traffic accidents in Morocco in order to generate an illustration of performance and safety. The intelligent approach and the statistical method were employed in this work to analyze statistical data on Moroccan road accidents. The statistical method employs data on the number of fatal accidents and the number of casualties in each deadly section. The factors affecting road infrastructure, on the other hand, were studied using ordinal regression. The intelligent methods refer to XGBoost and Bagging ELM (B-ELM). The B-ELM learns from the road infrastructure to identify the severity class of road sections, while XGBoost learns WSI weights. Figure 1 provides a step-by-step research methodology.

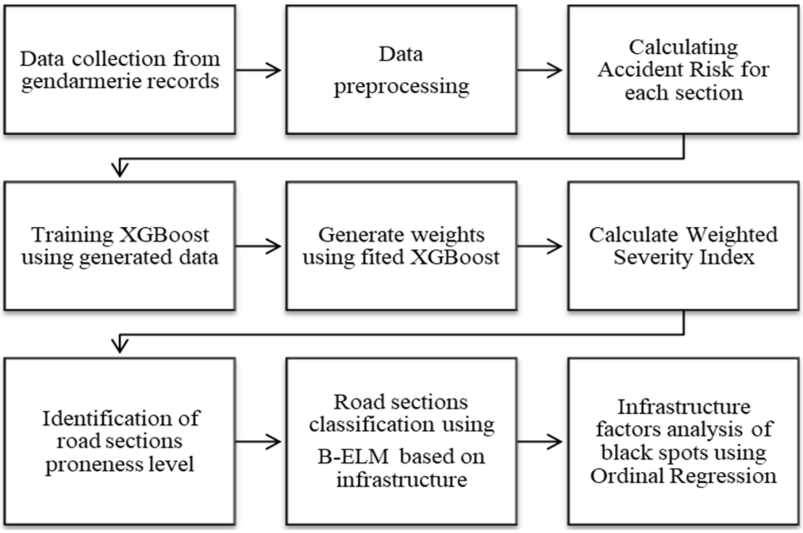

Figure 1. Research methodology

2.1. Data preprocessing

The data used in this study are obtained from the General Directorate of Roads, attached to the Ministry of Equipment and Water, Morocco. This directorate collects a large amount of data on traffic accidents on the entire road network. The datasets under consideration are made up of many separate files that provide statistics on traffic incidents that occurred between 2016 and 2017. Each file contains information on the accidents, as well as the drivers, passengers, and pedestrians that were involved. The metadata of the files includes temporal, spatial, vehicle, and road user information.

The data needs to be processed in order to derive insights. To begin, we concatenated and merged the individual tables into one that contains all the properties relevant to the study. Specifically, for each accident, spatial variables are collected, such as road number and kilometer point, as well as information about the road condition and the surrounding environment. Similarly, statistical variables such as the number of fatalities and injuries were collected. Investigations such as the one conducted by [33], [34] revealed that methods that apply the frequency of serious and fatal accidents perform better than those that apply the total accident frequency. They came to emphasize that accident severity is the key point in the process of assessing road accident black spots.

Accident fatality statistics are divided into four categories based on the time of death, namely, deaths on the scene, during transfer to hospital, 7 days, or 30 days following the accident. Injuries are divided into two groups according to severity, minor and severe injuries. The study considered 30,576 observations resulting in 3,604 deaths and 54,359 injuries. The spatial variables describe the road number, kilometer point, road condition, and other environmental characteristics surrounding the section in question. The road number factor describes 4 types of roads: national, regional, provincial, and unclassified. The data is then grouped by road number so that each could be analyzed independently of the others. We ended up with 652 different roads. We then divided each road into one-kilometer-long equidistant segments. The choice of a fixed distance has been shown to perform well on high-speed roads [2]. One advantage of the fixed-distance analysis for high-speed roads is that it avoids the problem of false positives (i.e., safe sections being included in the black spot study) or false negatives (i.e., road sections not being included in the black spot study when they are needed). Finally, the relevant factors and their descriptions are provided in Table 1. Additional details on road and victim characteristics are described in [14].

Table 1. Road features description

Factors	Description	Values
NUM ROU	Road number	Integer
PKM	Kilometer point of the accident location	Continuous
CASUALTIES	Statistics for two types of deaths and injuries	Death at the scene, death outside the scene, minor injury, serious injuries
ETA CHA	Road pavement condition	Good, bad, very bad, unpaved road
ETA SUR	Surface condition	Dry; wet, snowy, or icy; humid road; salted or sanded
LAR CHA	Road pavement width	Continuous in meters
TYP CHA	Road pavement type	Full road, Separated by a median
LAR ACO DRO (resp. LAR ACO GAU)	Right shoulder width (resp. Left shoulder width)	Continuous from 0 to 10 meters
POS ACO DRO (resp. POS ACO GAU)	Right shoulder position (resp. Left shoulder position)	At the same level as the pavement, not at the same level as the pavement
TOP	Road topography	Bridge, narrowing, obstacle
PRO LON	Length profile of the road	Flat, sloping, summit of a hill
TRA PLA	Road plan layout	Straight, left-hand curve, right-hand curve, double bend

The accident risk indicator (AR), which is the output variable, is then added to quantify the severity in terms of loss of life. It is intended to calculate the frequency of fatalities relative to the number of accidents in each section. The formula for the AR is as:

$$AR = \frac{F}{A} \tag{1}$$

where F denotes the number of fatalities and A denotes the number of accidents in the section. Road sections with a higher AR are considered more hazardous in terms of fatalities, meaning that there are significantly more fatalities per accident on these sections than on others of the same type, and therefore they need additional safety improvements.

2.2. XGBoost

Gradient boosting machines (GBM) are one of the best-performing algorithms for supervised learning, and XGBoost is one of the implementations of this technique [35]. It can be used to solve problems involving regression and classification. XGBoost is popular due to its fast out-of-core compute execution speed. Moreover, XGBoost provides critical functionality for calculating the contribution of the model's factors, known as the gain. Gain denotes the relative contribution of the relevant feature to the model, which is computed by adding each feature's contribution for each tree in the model. When this measure has a higher value than another characteristic, it means that the feature is more important in creating a forecast. Therefore, the most important factor to consider when determining the relative relevance of each feature is gain [36].

The XGBoost model was used to model the casualties' characteristics and Accident Risk severity rate. By using the four types of victims as input and the accident risk (AR) severity index as output, the hyperparameters of the XGBoost model were tuned to achieve an acceptable fit. Following the learning process, we calculated the score for each victim using the feature importance tool. The score of each feature indicates how important it is in the construction of the trees. The XGBoost model simulates the severity of the accident rate, yielding a set of input scores to be used in the WSI formula. It is worth noting that the scores are relative to the targets set.

2.3. Weighted severity index

One of the most commonly used techniques for locating black spots on a roadway is the weighted severity index or WSI. This method entails weighting bodily accidents in such a way that the more serious accidents are given a higher weight. In this study, accident casualties were categorized into four groups in the WSI method: instantaneous death, death from injuries, serious injury, and minor injury. From one approach to another, the weighting strategy is chosen differently. Under this study, each group was assigned a weight based on the analysis of AR using XGBoost, as described in the previous section.

The WSI formula is:

$$WSI(j) = \sum_{i=1}^{4} W_i \times A_i \quad (2)$$

where $WSI(j)$ is the WSI of the j^{th} stretch, W_i is the weight of the ith type of casualty, and A_i is the ith type of casualty.

The WSI was supposed to be a direct representation of accident risk. To put it another way, the higher the WSI of a section, the greater the chance of an accident. To identify the severity level of black spot sections, the mean and standard deviation of the WSI were computed in the second step of the study. The following three categories of accident black spot areas were determined using the WSI, mean, and standard deviation values as shown in Table 2. The concept is simple, if the WSI is below a specific limit, the expected number of fatal accidents at the spot is relatively small, therefore the location can be regarded as quite safe. But if the WSI exceeds a specific threshold, the likelihood of fatal accidents at the spot is greatly increased, and the location might be classified as excessively risky [2].

Table 2. Black spots classification based on the mean and standard deviation of WSI

Accident Severity Level	WSI
High	$WSI \geq Mean + 2\,S.D$
Medium	$Mean + 2\,S.D > WSI \geq Mean + 1.5\,S.D$
Low	$Mean + 1.5\,S.D > WSI \geq Mean + S.D$

2.4. Extreme learning machine

The extreme learning machine (ELM) is an innovative and fast SLFN learning algorithm in which the input weights and hidden layer biases are selected randomly. When learning the parameters of an ELM model, the optimization problem can be solved analytically, resulting in a closed form comprising just matrix multiplication and inversion. As a result, the learning process can be completed quickly without the use of an iterative method like a back-propagation neural network or the solution of a quadratic programming problem, as in the traditional SVM formulation. The input weights (connecting the input layer to the hidden layer) and hidden biases are randomly chosen in the ELM model, whereas the output weights (connecting the hidden layer to the output layer) are calculated using the Moore-Penrose approach.

In this study, we propose a Bagging ELM classifier model as shown in Figure 2 for predicting the state of road sections, i.e., white or black spots. There are four severity levels: safe, low, medium, and high, and the prediction model is based on road feature data. The size of the ELM hidden layer is carefully selected; we evaluated values ranging from 10 to 400 with a 10-step increment. The results of the evolution of model performance as a function of hidden neuron numbers demonstrate that the ideal number of hidden neurons with accurate results for our training dataset is 250 neurons. This approach has been detailed in a recent study [7]. For the purpose of Bagging base estimators, the number of ELM estimators was set to 10 since it shows better accuracy and requires less computation time. Bagging ELM consists in fitting several ELM models on different bootstrap samples and building a final predictor. The final prediction uses a voting approach based on a majority vote.

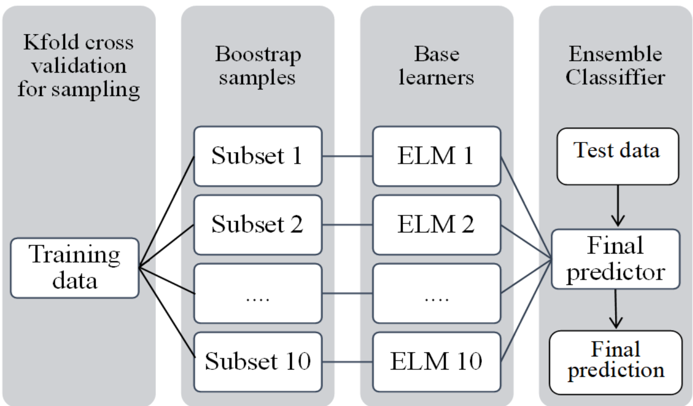

Figure 2. Bagging ELM explained

3. RESULTS AND DISCUSSION

This section summarizes and discusses the main findings of the work. We involved multiple stages to anticipate the state of road sections, and at each step, we obtained results that we have detailed below. The simulation results were obtained using a 2.00 GHz Intel Core i7-8550U computer using Python 3.8.0.

3.1. Determination of WSI weights

The XGBoost model was used with casualties as input and AR as output. To obtain an accurate model, the hyperparameters of XGBoost were tuned. Then the feature importance was explicitly estimated for each feature in the data, allowing features to be ordered. This gives us the casualty weights that will be used by WSI. Table 3 shows the results obtained from the preliminary analysis of XGBoost feature importance. As shown in the table, the shorter the interval between the accident and the mortality, the higher the score allocated. Furthermore, the score for those with serious injuries is higher than for those with minor injuries.

Table 3. Example of feature weights obtained by the XGBoost model

Death in spot	Death from injuries	Serious injuries	Minor injuries
21.88	11.79	5.45	3.15

3.2. Weighted severity index results

We calculated the WSI for each portion of the road using the total number of weighted characteristics; the WSI results are strictly classified between 0 and 100. Sections of the road with a higher WSI score will be considered high-level severity, while those with a lower value will be deemed more secure. Generally, WSI is a severity-based method used to identify accident black spots; the higher the WSI, the more severe the section. As a result, if a portion's WSI exceeds a specific threshold, the section is classed as a black spot. After the calculation of WSI, the mean and the standard deviation were computed to determine the severity level of each section. The identification results found 506 sections with high severity, 692 sections with medium severity, 760 sections with low severity, and 9,744 sections that represent safe spots. In general, the clustering of road sections points to a clearly unbalanced problem with significantly unbalanced groups. This result is expected since road accidents are unpredictable and discrete in nature.

3.3. Black spot classification

In classification problems, the fundamental disadvantage of unbalanced data is that they can lead to misleading results and even training failures. There are several approaches in supervised learning to reduce the impact of unbalanced data. Data resampling, cost-sensitive methods, and ensemble methods are the three primary categories of these techniques. The performance of the ELM model was evaluated on multi-class road accident data. Table 4 shows the available data, features, and classes. The data for the training set and the test set were created randomly, of which 20% were assigned to the test set.

Table 4. Final dataset properties

Dataset size	Number of features	Number of classes
11702	11	4

The accuracy results of the preliminary accident black spot classification are presented in Table 5. As shown in the table, the proposed Bagging ELM model outperforms the other models, Simple ELM, SVM, and MLP in terms of generalization. Moreover, it is fast in execution showing 0.27 seconds in the training phase, while SVM takes about 1.14 seconds and MLP 1.85 seconds. Thus, the execution speed was improved by 24% and 15% respectively. Obviously, the bagging ELM model outperformed the SVM and MLP models in terms of training and test accuracy. It was also able to provide substantial improvements in accuracy over the simple ELM.

Table 5. Performance results of B-ELM, ELM, SVM, and MLP

Model	Time (s)	Training accuracy	Test accuracy
ELM	**0.03**	81.6	90
Bagging ELM	0.27	**97.4**	**98.6**
SVM	1.14	80.9	77.2
MLP	1.85	83.6	89.8

3.4. Black spot analysis

Ananth [37] claim that ordinal models are underutilized in the social sciences because ordinal data are frequently reduced to a sequence of binary logistic regressions. Ordinal regression models, on the other hand, take advantage of the ordinality of the outcome by combining the relationships between the explanatory variables and the outcome into a single model. Unlike linear regression, which models a continuous dependent variable, ordinal regression (OR) models an ordered outcome variable. An ordinal variable contains information on the data's rank order. Ordinal regression is a popular method for modeling dependent variables with a natural ordering. The black spot dataset studies factors that influence whether a section of road is a black spot or not. Road sections were divided into four categories. Therefore, our outcome variable has four classes. The black spot (BS) severity level variable has a natural order, with the lowest order being the least severe sections and the greatest order being fatal sections. The severity of traffic accidents is categorized into four groups, with their equivalent class number: safe locations (0), low severity level (1), medium severity level (2), and high severity level (3).

The results of the odds ratio calculations and their 95% confidence intervals (CI) (lower; higher) from the parameter estimates are shown in Table 6. Quantitative factors, with positive coefficient values (Estimates), suggest that as the value increases, the probability of belonging to the black spot increases. Quantitative factors, with positive (resp. negative) coefficient values, suggest that if a section is part of the factor, i.e., if the section is characterized by the factor, it is more (resp. less) likely to belong to the black spot. They show that there is a significant effect of road condition and infrastructure on the BS severity outcome, namely plan layout, topography, right shoulder position, right shoulder width, and pavement width play an important role in the determination of the severity level of BS (Sig<0.05). According to the plan layout factor results, straight road has 7.684 times (95% CI; 2.786; 21.194) greater chance of having an effect in BS than double bend, Wald $\chi 2$=15,52, p<0.001. The odds of the left-hand curve affecting BS severity is 6.825 (95% CI; 2.458; 18.947) times more than the double bend, Wald $\chi 2$ =13,59, p<0.001.

Table 6. Results of ordinal regression model

Factors		Estimation	Wald	Sig	Odds	Lower	Upper
Pavement width		0,044	43,836	0,000	1,045	1,032	1,059
Right shoulder width		0,119	4	0,036	1,126	1,008	1,259
Position of the right shoulder	in level with pavement	1,014	9	0,003	2,756	1,400	5,426
	Not level with pavement	Base variable	-	-	-	-	-
Topography	Narrowing	-1,641	339	0,000	0,194	0,163	0,231
	Bridge	-0,215	6	0,013	0,807	0,681	0,955
	Obstacle	Base variable	-	-	-	-	-
Plan layout	Straight road	2,039	16	0,000	7,684	2,786	21,194
	Left-hand curve	1,921	14	0,000	6,825	2,458	18,947
	Right-hand curve	0,901	2,452	0,117	2,461	0,797	7,597
	Double bend	Base variable	-	-	-	-	-

There is also a strong association between the topography group and BS severity, even after other factors have been controlled. We can see significant and negative coefficients for bridge and narrowing. In fact, the odds of a narrowing section affecting BS severity is 0.194 (95% CI; 0.163; 0.231) times more than an obstacle, Wald $\chi 2$ =338,88, p<0.001. The odds of a bridge to have an effect in BS severity is 0.807 (95% CI; 0.681; 0.955) times than an obstacle, Wald $\chi 2$ = 6,19, p=0.013. Interestingly, there is also an effect of the right shoulder position, since the odds of the right shoulder not on level with the pavement having an effect on BS severity is 2.756 (95% CI; 1.4; 5.426) times higher than the right shoulder on level with the pavement, Wald $\chi 2$ =8,6, p=0.003. Also, the OR model indicates that right shoulder width and pavement width have a positive correlation with fatal accidents. More precisely, an increase in the right shoulder width (in meters) was associated with an increase in the odds of having severe BS with an odds ratio of 1.126 (95% CI; 1.008; 1.259), Wald $\chi 2$ (1)=4.386, p<0.05. Moreover, an increase in the pavement width (in meters) was associated with an increase in the odds of having a high-level severity of BS with an odds ratio of 1.045 (95% CI; 1.032; 1.059), Wald $\chi 2$ (1)=43.836, p<0.001. The coefficients of the pavement width and the right shoulder width are positive which suggests that as they increase, the likelihood of BS occurrence will increase. This indicates that with increasing pavement and right shoulder widths, accidents are prone to be more fatal, i.e., the severity level of black spots increases. On the other hand, the coefficients of narrowing and bridge are negatives, which suggests that if sections are narrowing or bridge, they are less likely to be a severe BS, i.e., accidents tend to be less fatal.

Obviously, we should not make this assumption without first checking the model's validity, particularly the proportional odds assumption. Looking at the model's fit, we can find a highly significant reduction in chi-square statistics (p.005), indicating that the model is clearly superior to the baseline or intercept-only model. According to the Nagelkerke R2, the model can account for 18% of the variation in BS severity level. The strategy of the road directorate for the identification of BS has resulted in 21 BS on the whole national road network, which are in agreement with the results of our classification as they present a high level of severity. All these sections are good and have a single pavement, have shoulders at the same level as the pavement, and are flat and straight. Furthermore, 95% have a dry surface, 48% are bridges and 38% are road narrowing. In addition, their width varies between 7 and 16 meters. In total, they caused 45 deaths, 216 serious injuries, and 520 accidents.

It was found like previous research [38]–[40] that curved and straight roads are the most likely to be black spots when considering the road layout, also the same results were obtained by Katre *et al.* [41] who found that the maximum number of accidents occurred on straight roads, slight curves, and sharp curves. Furthermore, the findings revealed that a one-meter increase in shoulder width was linked to an increase in BS severity level. The results are similar to the findings of [42] which demonstrate the positive effect of shoulder width on crash occurrence, and also by [43] which also found a positive correlation between crash severity and shoulder width. Indeed, an increase in the width of the shoulder, up to 2 meters, was found to increase the risk of accidents. The study also found that an increase in pavement width of one-meter leads to an increase in the severity level of black spots. This finding is consistent with previous research that reported that roads with a wider pavement width were generally associated with more severe crashes [44]–[46], also there is an agreement with Dadashova *et al.* [47] findings which showed that horizontal and vertical curvature and lane width are found to be associated with crash severity.

In contrast, the road length profile was identified as a non-significant factor in the severity level of black spots. These results reflect those of [38] who also found that road slopes did not have a significant impact on fatal crashes. Another important finding was the significant effect of shoulder width on black spot severity, this finding was also reported by Yasmin *et al.* [46] who highlighted the effect of lane width, and shoulder width on crash proportion by severity. The experiment detects evidence of the effect of right shoulder position on black spot severity, which is also consistent with the results of [44] which pointed out that shoulder elevation has an impact in determining black spot locations. Concerning the topography of the road section, results show that sections with obstacles reducing visibility are 5 times more likely to be severe locations than Narrowing of the Road and more than 1 time than Bridges. Consistent with the present results, previous studies have demonstrated the same thing, such as Bhatti *et al.* [48] investigated spatial factors associated with road crash locations. They found statistically significant associations between fatality risk and roadside obstacles located within 4 m of the roadside. One unanticipated finding was that there was no significant effect of left shoulder on black spot locations, this result reflects that of [49] which also found that the most influential variable for the presence of safe road sections is left shoulder width. Taken together, these results suggest that there is an association between road conditions and environmental features. Consequently, with road improvement road users can adapt their behaviors to new road conditions, which will lead to a positive impact on road safety.

4. CONCLUSION

This study proposes a model for the identification, classification, and analysis of black spots on rural roads of Morocco. The techniques used are weighted severity index (WSI), bagging extreme learning machine (B-ELM), and ordinal regression (OR). The model combines well-known methods and is based on spatial features such as road conditions and the environment. The WSI technique, along with, “XGBoost's feature importance” is used to calculate the severity of road sections. The B-ELM model for classification shows the best generalization in terms of accuracy, outperforming reputable models by reaching an accuracy of 98.6%. The regression results show that pavement width, right shoulder width and position, topography, and plan layout are risk factors for accident black spots.

The proposed model can help specialists in the field of road safety to make good decisions and improvements on high-priority road sections. In future work, we plan to prove the full potential of our approach by studying the entire Moroccan road network over a long period to extract the common properties of black spots. This method could further be refined by also including meteorological and temporal variables to improve the prediction accuracy. This research will make a substantial contribution to strengthening the intelligent transportation system in the country.

ACKNOWLEDGEMENTS

This document would not have been possible without the support of The General Directorate of Roads, which belongs to the Ministry of Equipment and Water. The center has made every effort to share with us the process they use to monitor, manage and process road traffic data and to provide us with the indispensable real data of the road accidents that we have analyzed in this study.

REFERENCES

[1] Ministry of Equipment and Water, “Accident statistics,” (in French) Ministry of Equipment and Water, Morroco, http://www.equipement.gov.ma/Infrastructures-Routieres/Reseau-Routier-du-Royaume/Pages/Statistiques-des-accidents.aspx (accessed May 04, 2020).

[2] M. Ghadi and Á. Török, “Comparison different black spot identification methods,” *Transportation Research Procedia*, vol. 27, pp. 1105–1112, 2017, doi: 10.1016/j.trpro.2017.12.104.

[3] H. Xin-min, L. Tao, Z. Tian-hua, H. Yong, X. Hai-wei, and Y. Tao, “Highway black spot recognition and improvement,” *Journal of Traffic and Transportation Engineering*, vol. 4, no. 1, pp. 106–109, 2004.

[4] W. Cheng and S. Washington, “New criteria for evaluating methods of identifying hot spots,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2083, no. 1, pp. 76–85, Jan. 2008, doi: 10.3141/2083-09.

[5] V. Ratanavaraha and D. Watthanaklang, “Road safety audit: identification of bus hazardous location in Thailand,” *Indian Journal of Science and Technology*, vol. 6, no. 8, pp. 5126–5133, 2013, doi: 10.17485/ijst/2013/v6i8.6.

[6] I. W. Agustin, “Analysis of car accident at the location of black-spot and rating for accident-prone roads in Surabaya,” *IOP Conference Series Earth and Environmental Science*, vol. 328, no. 1, Oct. 2019, doi: 10.1088/1755-1315/328/1/012023.

[7] 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*, vol. 11, no. 12, Nov. 2020, doi: 10.3390/info11120542.

[8] L. Breiman, “Bagging predictors,” *Machine Learning*, vol. 24, no. 2, pp. 123–140, Aug. 1996, doi: 10.1007/BF00058655.

[9] Z.-H. Zhou, “Ensemble learning: foundations and algorithms,” Taylor and Francis: Boca Raton, FL, USA, 2012.

[10] G. Huang, H. Zhou, X. Ding, and R. Zhang, “Extreme learning machine for regression and multiclass classification,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 2, pp. 513–529, Apr. 2012, doi: 10.1109/TSMCB.2011.2168604.

[11] M. Ghadi and Á. Török, “A comparative analysis of black spot identification methods and road accident segmentation methods,” *Accident Analysis and Prevention*, vol. 128, pp. 1–7, Jul. 2019, doi: 10.1016/j.aap.2019.03.002.

[12] N. Srinivasan, V. S. Iyer, M. Chand, and K. Srinath, “Scientific identification and improvement of accident prone locations on national highways in Kerala,” *Journal of the Indian Roads Congress*, vol. 48, no. 3, pp. 477–525, 1987.

[13] T. Sayed, W. Abdelwahab, and F. Navin, “Identifying accident-prone locations using fuzzy pattern recognition,” *Journal of Transportation Engineering*, vol. 121, no. 4, pp. 352–358, Jul. 1995, doi: 10.1061/(ASCE)0733-947X(1995)121:4(352).

[14] 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)*, Jun. 2020, pp. 1–7, doi: 10.1109/ISCV49265.2020.9204325.

[15] I. El Manaa, F. Benjelloun, M. A. Sabri, A. Yahyaouy, and A. Aarab, “Road traffic anomaly detection: a survey,” in *Digital Technologies and Applications*, Springer International Publishing, 2022, pp. 772–781.

[16] C. Zhang, Y. Shu, and L. Yan, “A novel identification model for road traffic accident black spots: a case study in Ningbo, China,” *IEEE Access*, vol. 7, pp. 140197–140205, 2019, doi: 10.1109/ACCESS.2019.2942647.

[17] M. Q. Ghadi and Á. Török, “Evaluation of the impact of spatial and environmental accident factors on severity patterns of road segments,” *Periodica Polytechnica Transportation Engineering*, vol. 49, no. 2, pp. 146–155, Apr. 2020, doi: 10.3311/PPtr.14692.

[18] A. Iqbal, Z. U. Rehman, S. Ali, K. Ullah, and U. Ghani, “Road traffic accident analysis and identification of black spot locations on highway,” *Civil Engineering Journal*, vol. 6, no. 12, pp. 2448–2456, Dec. 2020, doi: 10.28991/cej-2020-03091629.

[19] F. Calsavara, F. I. Kabbach, and A. P. C. Larocca, “Effects of fog in a Brazilian road segment analyzed by a driving simulator for sustainable transport: drivers’ speed profile under in-vehicle warning systems,” *Sustainability*, vol. 13, no. 19, Sep. 2021, doi: 10.3390/su131910501.

[20] A. Mbarek, M. Jiber, A. Yahyaouy, and A. Sabri, “A new model for black spots identification using weighted severity index,” in *2021 Fifth International Conference on Intelligent Computing in Data Sciences (ICDS)*, Oct. 2021, pp. 1–7, doi: 10.1109/ICDS53782.2021.9626758.

[21] R. D. Abdualjabar and O. A. Awad, “Parallel extreme gradient boosting classifier for lung cancer detection,” *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 24, no. 3, pp. 1610–1617, Dec. 2021, doi: 10.11591/ijeecs.v24.i3.pp1610-1617.

[22] M. Panigrahi, D. K. Behera, and K. C. Patra, “Epileptic seizure classification of electroencephalogram signals using extreme gradient boosting classifier,” *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 25, no. 2, pp. 884–891, Feb. 2022, doi: 10.11591/ijeecs.v25.i2.pp884-891.

[23] T. A. Assegie, T. Karpagam, R. Mothukuri, R. L. Tulasi, and M. F. Engidaye, “Extraction of human understandable insight from machine learning model for diabetes prediction,” *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 11, no. 2, pp. 1126–1133, Apr. 2022, doi: 10.11591/eei.v11i2.3391.

[24] E. Sugiharti, R. Arifudin, D. T. Wiyanti, and A. B. Susilo, “Integration of convolutional neural network and extreme gradient boosting for breast cancer detection,” *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 11, no. 2, pp. 803–813, Apr. 2022, doi: 10.11591/eei.v11i2.3562.

[25] M. A. Azim, M. K. Islam, M. M. Rahman, and F. Jahan, “An effective feature extraction method for rice leaf disease classification,” *Telecommunication Computing Electronics and Control (TELKOMNIKA)*, vol. 19, no. 2, Apr. 2021, doi: 10.12928/telkomnika.v19i2.16488.

[26] M. Vasudevan, R. S. Narayanan, S. F. Nakeeb, and A. Abhishek, “Customer churn analysis using XGBoosted decision trees,” *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 25, no. 1, pp. 488–495, Jan. 2022, doi: 10.11591/ijeecs.v25.i1.pp488-495.

[27] D.-D. Nguyen, M.-T. Le, and T.-L. Cung, “Improving intrusion detection in SCADA systems using stacking ensemble of tree-based models,” *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 11, no. 1, pp. 119–127, Feb. 2022, doi: 10.11591/eei.v11i1.3334.

[28] B. Debrabant, U. Halekoh, W. H. Bonat, D. L. Hansen, J. Hjelmborg, and J. Lauritsen, “Identifying traffic accident black spots with Poisson-Tweedie models,” *Accident Analysis and Prevention*, vol. 111, pp. 147–154, Feb. 2018, doi: 10.1016/j.aap.2017.11.021.

[29] D. Saha, P. Alluri, E. Dumbaugh, and A. Gan, “Application of the Poisson-Tweedie distribution in analyzing crash frequency data,” *Accident Analysis and Prevention*, vol. 137, Mar. 2020, doi: 10.1016/j.aap.2020.105456.

[30] H. H. Nguyen, P. Taneerananon, and P. Luathep, “Approach to identifying black spots based on potential saving in accident costs,” *Engineering Journal*, vol. 20, no. 2, pp. 109–122, May 2016, doi: 10.4186/ej.2016.20.2.109.

[31] P. Rvyas, M. L. Honnappanavar, and H. Balakrishna, "Identification of black spots for safe commuting using weighted severity index and GIS," *people*, vol. 31, 2012.
[32] H. Zhu, Y. Zhou, and Y. Chen, "Identification of potential traffic accident hot spots based on accident data and GIS," *MATEC Web of Conferences*, vol. 325, Oct. 2020, doi: 10.1051/matecconf/202032501005.
[33] A. Montella, "A comparative analysis of hotspot identification methods," *Accident Analysis and Prevention*, vol. 42, no. 2, pp. 571–581, Mar. 2010, doi: 10.1016/j.aap.2009.09.025.
[34] X. Qu and Q. Meng, "A note on hotspot identification for urban expressways," *Safety Science*, vol. 66, pp. 87–91, Jul. 2014, doi: 10.1016/j.ssci.2014.02.006.
[35] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794.
[36] "XGBoost Python package," XGBoost. https://xgboost.readthedocs.io/en/latest/python/index.html (accessed May 21, 2022).
[37] C. Ananth, "Regression models for ordinal responses: a review of methods and applications," *International Journal of Epidemiology*, vol. 26, no. 6, pp. 1323–1333, Dec. 1997, doi: 10.1093/ije/26.6.1323.
[38] X. Liu and J. (Cecilia) 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.
[39] S. Reed and A. Morris, "Characteristics of fatal single-vehicle crashes in Europe," *International Journal of Crashworthiness*, vol. 17, no. 6, pp. 655–664, Dec. 2012, doi: 10.1080/13588265.2012.709394.
[40] S. Barman and R. Bandyopadhyaya, "Crash severity analysis for low-speed roads using structural equation modeling considering shoulder- and pavement-distress conditions," *Journal of Transportation Engineering, Part A: Systems*, vol. 146, no. 7, Jul. 2020, doi: 10.1061/JTEPBS.0000373.
[41] N. Katre, N. Pitale, and S. Bobade, "Analysis of black spots on NH-3 and its rectification," *Journal of Transportation Systems*, vol. 4, no. 2, 2019.
[42] N. Stamatiadis, J. G. Pigman, D. Lord, J. Sacksteder, and W. Ruff, *Impact of shoulder width and median width on safety*. Washington, D.C.: National Academies Press, 2009.
[43] V. Gitelman, E. Doveh, R. Carmel, and S. Hakkert, "The influence of shoulder characteristics on the safety level of two-lane roads: A case-study," *Accident Analysis and Prevention*, vol. 122, pp. 108–118, Jan. 2019, doi: 10.1016/j.aap.2018.10.003.
[44] G. Sugiyanto, Y. A. Prasetyo, E. W. Indriyati, Y. Yanto, and M. Y. Santi, "Determining of black spot location in Purbalingga Regency using road geometric approach," *IOP Conference Series: Earth and Environmental Science*, vol. 847, no. 1, Sep. 2021, doi: 10.1088/1755-1315/847/1/012019.
[45] S. Vadlamani, E. Chen, S. Ahn, and S. Washington, "Identifying large truck hot spots using crash counts and PDOEs," *Journal of Transportation Engineering*, vol. 137, no. 1, pp. 11–21, Jan. 2011, doi: 10.1061/(ASCE)TE.1943-5436.0000183.
[46] S. Yasmin, N. Eluru, J. Lee, and M. Abdel-Aty, "Ordered fractional split approach for aggregate injury severity modeling," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2583, no. 1, pp. 119–126, Jan. 2016, doi: 10.3141/2583-15.
[47] B. Dadashova, C. Silvestri-Dobrovolny, J. Chauhan, M. Perez, and R. Bligh, "Hot-spot analysis of motorcyclist crashes involving fixed objects using multinomial logit and data mining tools," *Journal of Transportation Safety and Security*, vol. 14, no. 7, pp. 1201–1219, Jul. 2022, doi: 10.1080/19439962.2021.1898070.
[48] J. A. Bhatti, J. Sobngwi-Tambekou, E. Lagarde, and L.-R. Salmi, "Situational factors associated with road traffic crashes: A case-control study on the Yaoundé-Douala road section, Cameroon," *International Journal of Injury Control and Safety Promotion*, vol. 17, no. 4, pp. 215–222, Dec. 2010, doi: 10.1080/17457301003728510.
[49] E. de la Peña González, "Design of a model for the identification and analysis of road sections without accidents: a new vision of road safety," (in Spanish), PhD Thesis, Universidad Politecnica de Madrid - University Library, 2015.

BIOGRAPHIES OF AUTHORS

Abdelilah Mbarek is a Ph.D. student at the Department of Computer Science 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 University of Sidi Mohamed Ben Abdellah, Fez, Morocco, His current research interests lie in machine learning, computer vision, and data analysis. He can be contacted at email: abdelilah.mbarek@hotmail.com.

Mouna Jiber received her Ph.D. in computer science from the USMBA-Fez University 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 lie in machine learning, intelligent transportation systems, and data analysis. She can be contacted at email: m.jiber@mtpnet.gov.ma.

Ali Yahyaouy received his Ph.D. in computer science from the USMBA-Fez University in 2010. He is currently a professor of computer science at FSDM-Fez. He is also a member of the LISAC Laboratory. His main research activities are concerning MAS, reinforcement learning, machine learning, traffic control, and energy management. He can be contacted at email: ali.yahyaouy@usmba.ac.ma.

Abdelouahed Sabri is a professor in the Computer Science Department at the Faculty of Sciences, Fez Morocco. Member of the LIIAN laboratory. Coordinator of the SIGL Professional License (Information Systems and Software Engineering) since 2014. He received his Ph.D. degree in 2009 from the Faculty of Sciences, Fez Morocco. His research areas include image processing and information retrieval. He can be contacted at email: abdelouahed.sabri@usmba.ac.ma.