IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 6, December 2025, pp. 4461~4473
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i6.pp4461-4473

Predicting the severity of road traffic accidents Morocco: a supervised machine learning approach

Halima Drissi Touzani[1], Sanaa Faquir[1,2], Ali Yahyaouy[1]
[1]LISAC Laboratory, Faculty of Science Dhar Mehraz, Sidi Mohamed Ben Abdellah University, Fes, Morocco
[2]Department of Engineering and Technology, Vaughn College of Aeronautics and Technology, New York, United States

Article Info

Article history:

Received Apr 3, 2025
Revised Aug 17, 2025
Accepted Sep 7, 2025

Keywords:

Accidents severity prediction
Data analytics
Human factors
Road accidents
Supervised machine learning

ABSTRACT

Early prediction of road accidents fatality and injuries severity is one of the important subjects to road safety emphasizing the critical need to prevent serious consequences to reduce injuries and fatalities. This study uses real road accidents data set in Morocco. It represents the intersection between road safety and data science, aiming to employ machine learning techniques to provide valuable insights in accident's severity prevention. The purpose of this paper is to study road accidents data in the country and combine results from statistical methods, spatial analysis, and machine learning models to determine which factors will mostly contribute to increase the accident' severity in the country. A comparison of results obtained was also conducted in this paper using different metrics to evaluate the effectiveness of each method and determine the most important factors that contribute to increase the fatality or injuries severity in the specific context of accidents. The best prediction model was then injected into a proposed algorithm where more intelligent techniques are included to be implemented in a car engine to perform an early detection of severe accidents and therefore preventing crashes from happening.

Corresponding Author:

Halima Drissi Touzani
LISAC Laboratory, Faculty of Science Dhar Mehraz, Sidi Mohamed Ben Abdellah University
Appt 5 imm 102 Errachidia 2 Route Ain Chkef, Fes, Morocco
Email: halima.drissitouzaniwalali@usmba.ac.ma

1. INTRODUCTION

A road accident is an event that involves one or more vehicles on a public road resulting in material damage related only to objects or physical damage that can be serious by involving at least one injured person or even fatal with at least one person killed. Every year, road safety officials set up awareness campaigns on this subject and public policies take new measures such as lowering the speed limit to 80 on roads. Despite this work, road fatalities remain too high all around the world. In France, the 2022 report on road accidents stated that 3,541 people were dead on the metropolitan roads of France or overseas. It was also reported that the number of fatal accidents increased by 1.3% compared to 2019 [1]. In the country of Slovakia, road traffic accidents have arisen as a critical public health matter, demanding a multidisciplinary approach for effective resolution. More than 40,000 death every year on the roads even when road fatalities are decreasing. Some preventions were applied such as in-vehicle safety and driver assistance systems to interfere before the accident happens [2]. According to Ennajih *et al.* [3], there has been a noticeable increase in road accident-related fatalities since 1968. The traffic accidents have become the leading cause of death among young people aged 17 to 29 years. Over 90% of deaths and injuries due to traffic accidents occur in low- and middle-income countries. During the last three years, traffic accidents have been increasing from September to December yearly in Taiwan Province of China according to big data analysis of historical accident data. The latest statistics from the road safety information platform of the Ministry of Transport showed a slight decrease in traffic accident deaths due to some safety measures taken after various accident prevention work [4]. Portugal also is facing a serious problem with fatal accidents. In 2020, the country revealed a total number of 27,725 accidents with victims from which 536 victims dead. These numbers put the country at the ninth-highest position with the most fatalities per million inhabitants in the European Union [5].

In the country of Morocco as well, car accidents became so unsignificant since there are people passing away every day. Based on statistics released by the Ministry of Equipment and Transport in 2012, there were 44,902 car accidents, resulting in the loss of 10 lives and leaving 140 injured per day. These incidents not only devastate entire families but also lead to the creation of orphans and widows [6]. In line with the data of the National Road Safety Agency (NARSA), the newest statistics show that Morocco recorded 3,499 road deaths in 2022, marking a 5% regression compared to 2021. Nearly three-quarters of these fatalities involve vulnerable road users [7]. Figure 1 shows that Morocco has the highest fatality risk 5.6 road deaths of per 100,000 registered vehicles in 2022 compared to other countries.

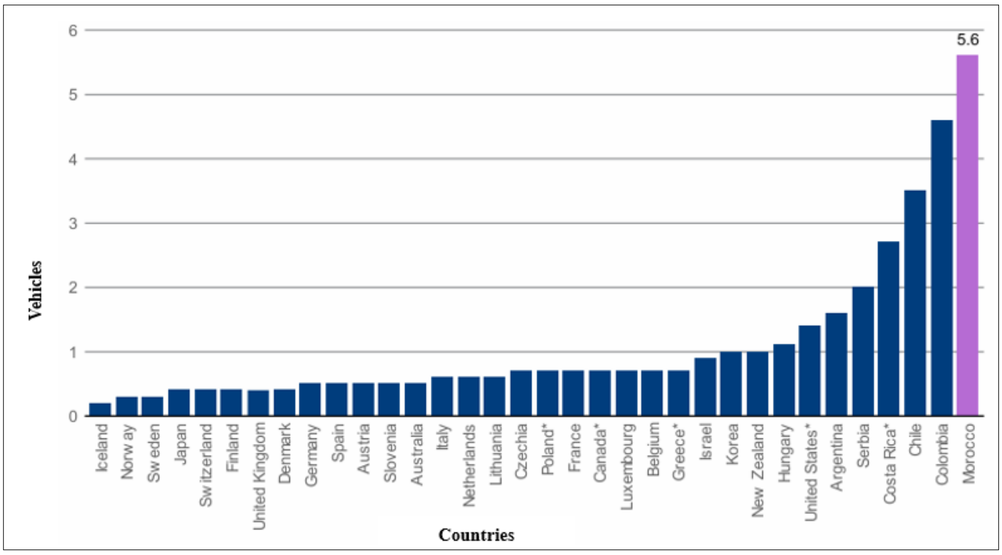


Figure 1. Road fatalities per 10,000 registered vehicles in Morocco compared to other countries in 2022

After all these alarming statistics, it is very important to apply different techniques for effective prevention that can help in reducing human losses and irreparable physical and psychological damage. However, building predictive models requires detecting factors responsible for the occurrence of these accidents and their dangerousness. Different predictive models have been presented in the literature. Results depend largely on the variables and type of the data sample analyzed. Different analytic methods and machine learning algorithms used to find various accident risk factors and develop predictions available were listed [8]. In general, most studies on traffic accident forecasting focused on two important research methods: statistical methods and neural networks. Statistical techniques such as k-nearest neighbors (KNN), support vector machine (SVM), and logistic regression (LR) were employed to forecast the frequency of the traffic accidents while artificial neural networks (ANN) were applied to implement the flexibility, generalization, or more robust prediction system [9]. Detecting accidents black spots was also an important issue to improve road traffic safety and reduce the traffic accidents severity. Various studies assessed the efficacy of the empirical Bayesian technique on accidents black spot classification as it was proposed in [10] to construct a black spot identification model that could identify a black spot. The model gave the best accuracy compared to the ID3 decision tree, LR, and SVM [11]. Classification algorithm such as KNN, naïve Bayes (NB), and ANN were utilized in [12] to develop a predictive model for predicting occurrences of traffic accidents. However, from all of this research, only a few used machine learning methods to predict car crashes' fatality or the seriousness of injuries in Turkey, injuries and fatalities data analysis were analyzed using non-linear regression and ANN [13].

To determine the most important factors leading to car accidents, data mining techniques such as decision trees, non-linear regression, and classification methods were used in [14] and results revealed the vehicle type factor as one of the most dangerous factors related to accident severity. A traffic crash risk prediction model using the limited random-synthetic minority over-sampling technique (LR-SMOTE) algorithm was used in [15] to categorize the sharp acceleration and deceleration, volume, average speed, the quotient between free flow speed and current average road speed and the coefficient of variation of speed as being the main attributes that affect the risk of a crash in an accident. The model used real-time traffic flow data and risky driving behavior data to study the traffic crash risk on freeways. In the UK as well, an in-depth analysis of the current ML models was implemented for predicting injury severity in road accidents [16].

The Haddon matrix was used in this context to analyze the causes and seriousness of the accidents in the country. Chakraborty *et al.* [17] conducted in India sought to develop a model for predicting fatal pedestrian crashes and to determine the factors that exacerbate the severity of transportation-related injuries and deaths. The research revealed that the 'approaching speed' of motorized vehicles was the most pronounced influence on fatal pedestrian crashes.

This paper includes two data mining techniques: data analytics and and supervised machine learning algorithms to predict the severity of accidents on Moroccan roads. An accident is considered a fatal accident if at least one person was killed. When no person is killed in the accident but there are injuries then the accident is not fatal. In this case the injuries seriousness is measured. The aim of this research is to determine a predictive model for forecasting in advance the fatality of injury severity of an accident to happen based on accidents heterogeneous characteristics. The study used real road traffic accidents provided by the NARSA for the two years 2015 and 2016. The prediction is obtained for each case given different accident attributes and combines results from statistical methods, spatial analysis, and artificial intelligence models. Statistics methods function well in the processing and data analysis, however, different studies demonstrated that ANNs give more accurate prediction capabilities [18], [19]. In this research, five different suprevised machine learning methods were used: NB, SVM, ANN, KNN, LR. The performance of these models was compared using different metrics on the provided datasets. The attributes in the dataset were grouped by three factors as described in the Haddon's matrix (human factor, vehicle, and equipment factor and environment factor). The random forest importance technique was also used as an embedded feature selection method in machine learning to identify the attributes that contribute more to accidents severity. At the end, a proposed algorithm was designed to include a collection of data from different devices, the best prediction model along with other intelligent techniques such as advanced driver-assistance systems (ADAS). This algorithm can be implemented in semi-autonomous vehicles to enable early detection and prevention of severe accidents.

2. METHOD

Road traffic accidents pose a significant challenge to road safety, emphasizing the critical need to prevent their serious consequences to reduce injuries and fatalities. In this context, leveraging machine learning models shows promise for predicting accident severity and facilitating swift intervention. Solutions and decisions can be found without data. However, these decisions are only based on personal heuristics developed by lived experience. To find more concrete decisions, especially in critical domains such as road traffic accidents requires a lot of data and thinking. Due to the development of internet of things (IoT) devices connected to the network, data collection and transmission became practical for further analysis. Having good data is essential for making the right decisions.

Thirty years ago in the USA, William Haddon Jr. described road transport as a poorly designed "man-machine" system requiring comprehensive systemic treatment [20]. The matrix Haddon's nine-cell model represents the dynamic system, with each cell offering intervention options to reduce traffic injury. The matrix has led to significant advances in understanding the behavioral, road and vehicle factors that influence the number and severity of injuries in traffic accidents. The aim was to achieve specific reductions in the number of road fatalities and injuries all around the world. However, the practical application of this systemic approach remains the main challenge for policymakers and road safety professionals. Our research focuses on finding solutions using machine learning techniques to prevent accidents before happening considering the factors listed in Haddon's matrix by defining an accident as a series of features where data analysis and prediction methods can be applied to help reducing the cost of fatalities and driver injuries.

2.1. Data cleaning and preprocessing

The dataset used contained two files. One file contains information on the accident's drivers, passengers involved as well as the accident outcome indicators (COD_TUE for fatalities and COD_BLE for injuries) while the other file includes temporal, spatial, vehicle, and road user information. These two files were merged into one complete dataset to be able to conduct the research. 94,862 accident cases were analysed after cleaning the database from incomplete information. The dataset holds essential information, offering a comprehensive understanding of the circumstances surrounding each accident: numerical, categorical, and temporal data. These attributes were classified based on the three factors as indicated in

Haddon’s matrix: human factors as shown in Table 1 such as driver’s personal information, excessive or inappropriate speeding, failure to respect safety distances, vehicles and equipment factors such as technical problems related to the condition of the vehicle and environmental factors like weather conditions, road conditions, and others as stated in details [21].

Table 1. Application of Haddon's matrix on the accidents data set

Phase		Factors Human	Vehicles and equipment	Environment
Before accident	Accident prevention	Driver’s age Driver’s sex Driver’s profession Driver’s type of driving license	Vehicle type Usage of the vehicle	Code province Pavement condition Surface condition Light Weather conditions
Accident	Trauma prevention in case of accident	Errors due to the driver, physical condition of the driver	Obstacles hurt	Road detour
After accident	Maintain alive	Killed/injured or no problem		

Upon initial examination, the datasets revealed multiple challenges. Several columns had a significant percentage of missing data. Other columns contained very few non-null values. These columns appeared largely uninformative. Additional columns were redundant while further ones appeared to be not relevant to the analysis goals, such as specific codes or identifiers with no related codification or explanation. The accident date format was also converted to datetime format to facilitate temporal manipulation. This will allow the model to exploit the temporal aspects of accidents, such as the variation in severity levels depending on the days of the week, months, or seasons.

All the challenges mentioned above required a thorough cleaning and preprocessing phase to ensure the reliability and accuracy of the subsequent analysis. This phase was performed by applying the filters, imputation methods and selection phases available by the RapidMiner which is a powerful data science platform that offers an integrated environment for data preparation, machine learning, deep learning, text mining, and predictive analytics. To be able to do the analysis phase, the two databases mentioned above were merged to one complete dataset using phyton. Python was chosen for its flexibility and the powerful data handling capabilities offered by its libraries. The NumPy and Pandas libraries were imported to use classification functions. The first function 'classify_injury_severity', categorizes injuries according to the severity code ('COD_BLE'). The second, 'classify_mortality', determines the overall status of the victim ('fatal or non-fatal') by checking whether there are any dead or injured people in the dataset. A third function, 'classify_injury_gravity', considers both fatalities and injuries to classify the severity of the injuries ('injury severity'). Results of the classification functions were added to the dataset as new columns, 'fatal or not' and 'injury severity'. These were the target attributes for this research. Python was used in tandem with RapidMiner for more complex data transformations and for tasks requiring custom scripting. In conclusion, data preprocessing involved the meticulous transformation, selection, and imputation of features, as well as the explicit definition of the target variable. These procedures were indispensable for guaranteeing data quality and the robustness of predictive models.

2.2. Descriptive statiscics

An initial data exploration to comprehend the scope of the available information was conducted on the same database in [22] using Excel graphs. The study revealed some hidden information such as the sex of the driver that is common in the accidents, the most likely timing when most accidents occur and other information that can’t be grabbed directly with the massive data presented. This exploration facilitated the identification of characteristics that may have relevance in predicting accident severity.

Before conducting the research, few terms should be explained: a person in an accident can be either killed or injured. If the accident has at least one person killed, then accident is categorized as a fatal accident (the value in the fatality column in the database is 1 if at least one person killed and 0 if no person was killed). A person can be killed immediately or dies within 30 days due after the car crash. In the case when no person is killed in the accident but there are injuries then the accident is not fatal. In this case the injuries severity is measured. This last is categorized into three classes: i) a person is seriously injured: any person injured in a road crash requiring hospitalization for six days or more (the value in the injury column in the database is 2 if at least one person is seriously injured); ii) a person slightly injured: any person injured in a road crash requiring medical treatment or hospitalization of fewer than six days (the value in the injury column in the database is 1 if no one is seriously injured and at least one person is slightly injured); and iii) there are no injuries at all in the accident (the value in the injury column in the database is 0 if no one is injured).

Using Excel as well, specific data analysis using different types of graphs was presented in Figure 2 to measure the severity of accidents. The graphs show that for the accident cases presented in this database 37% of people are killed in the accidents which is an important number in two years. It also shows that for non-fatal accidents, almost 50% of people involved in accidents are injured with 12% seriously injured.

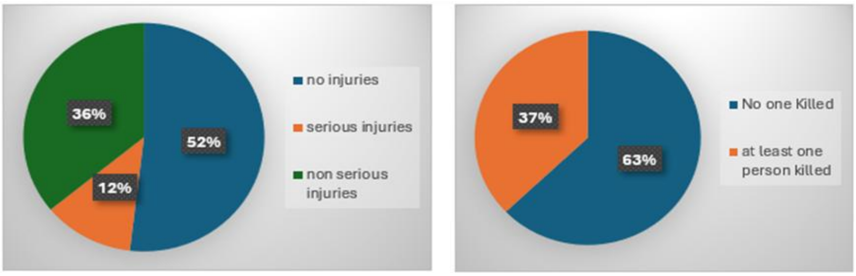


Figure 2. Accident severity and fatality percentage for 2015-2016

In addition to the primary data fields, the database contains detailed mappings for a multitude of coded values, offering descriptions for various conditions and categories, such as driver's license type, vehicle categories, occupations, and injury types, among others. This allows for a detailed and nuanced analysis of accidents. The additional attributes added, such as the total number of fatalities and injuries as well as the estimated severity of injuries, further enrich the data set. These fields allow for an accurate assessment of the impact of each accident, facilitating advanced statistical analyses to determine the most significant factors contributing to road incidents.

Heatmaps are other ways to easily visualize and analyze complex data. The heatmap in Figure 3 displays the average number of people involved in the accident for different hours of the day and different days of the week. The intensity of the color represents the magnitude, with darker colors indicating likely higher averages. The heatmap was implemented using RapidMiner tool and it shows that there is a relationship between the number of accidents that happen during different days over time. By observing how cell colors change across each axis, we can notice that most accidents occur during nighttime until almost 9 am. This can be explained by the less illumination at night and the traffic timing from 7 to 9 am. The color is less dark between 12 and 13 pm on weekdays. This can be explained by the traffic during lunchtime as well.

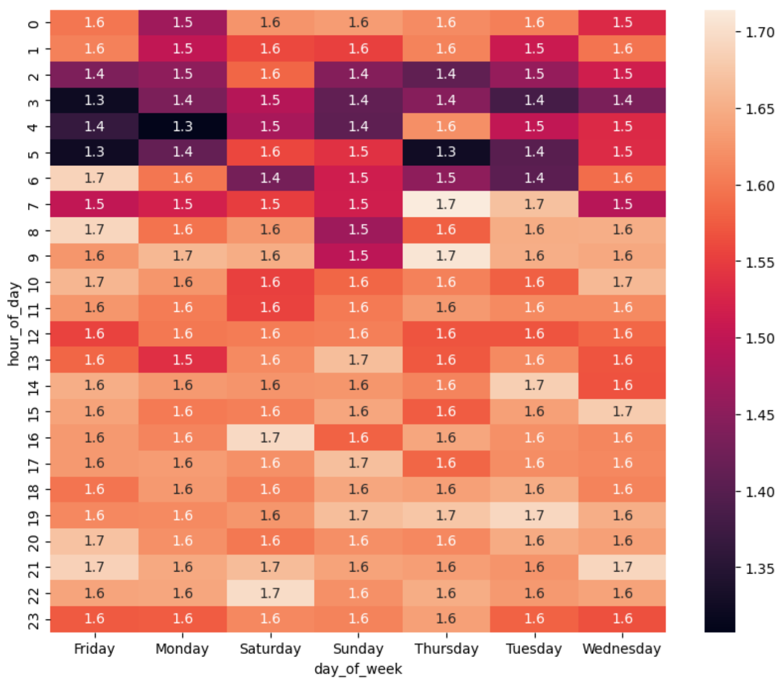


Figure 2. Heatmap of daily accidents by hours distribution

The correlation matrix of variables was also implemented as shown in Figure 4. This matrix shows the correlation coefficients between pairs of variables in our dataset. Darker colors (red or violet) indicate stronger positive or negative correlations. Some variables, such as killed_instantly, non_killed, seriously_injured, and lightly_injured, show higher correlations among themselves as expected.

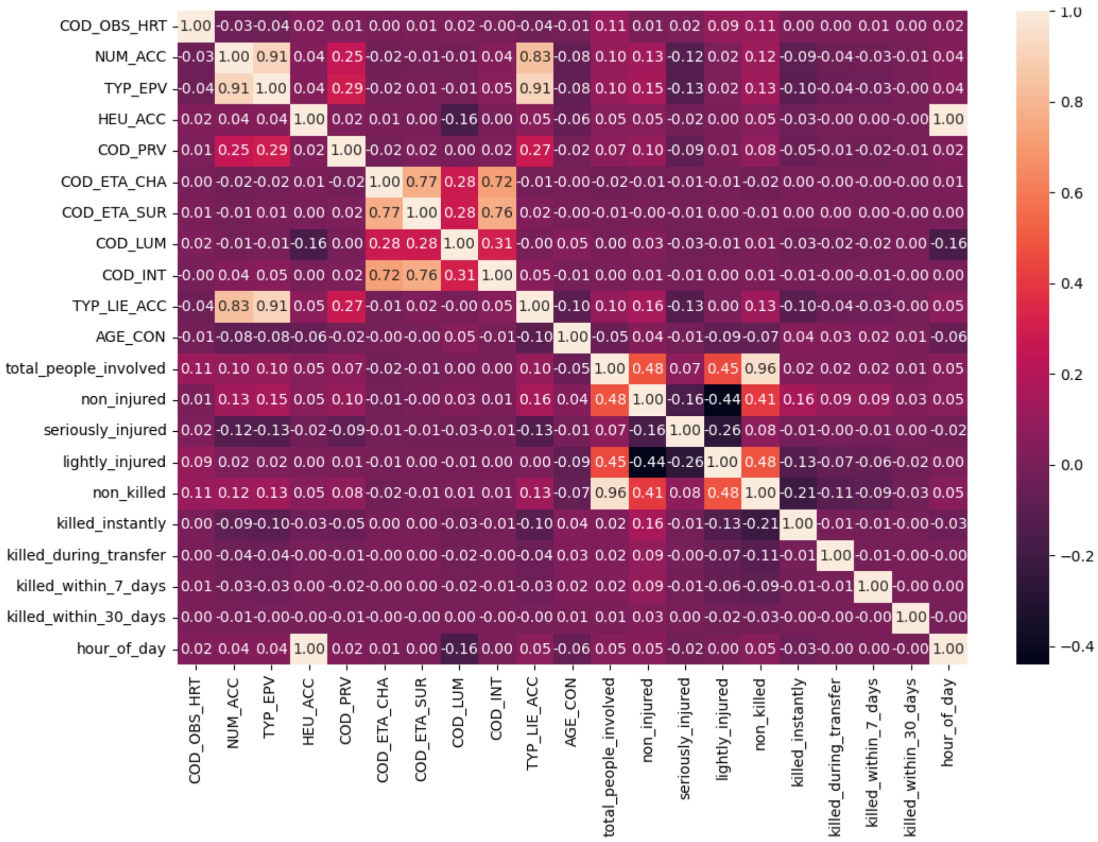


Figure 4. Matrix of correlation of variables

2.3. Supervised machine learning algorithms

In previous work, unsupervised machine learning methods were applied on a different real road traffic accidents dataset to gain valuable insights into accident patterns and trends. Results showed that accidents could be categorized by day and night based on four attributes: type of collision, initial shock, and the movement in the accident. Valuable suggestions were then sent to the Ministry of Transport to help reduce road accidents [23] and injected into a fuzzy logic controller to train a semi-autonomous car to take the right decision when the driver doesn't react timely and properly [24]. However, since the country of Morocco had a high cost of road crashes in 2022 estimated at EUR 1.6 billion without counting the cost of slight injuries and property damage [7] as shown in Table 2, more data analytics using supervised machine learning algorithms was conducted in this research to perform an early predict of accidents severity and therefore help reduce this cost in the country.

Table 2. Cost of road crashes in Morocco, 2022

Fatalities	Unit cost (EUR)	Number	Total cost (EUR)
	256136	3499	0.9 billion
Seriously injured	64033	10929	0.7 billion
Total			1.6 billion

The dataset used in this research contains various information such as accident time, location, event type, and brightness which provide crucial insights into accident circumstances for the model development.

The target variable, fatalities, or injuries was selected to measure the severity of the accident. In case the accident is not fatal, another target variable is chosen to represent a categorization of injury severity in the accident. Different machine learning models to predict the fatality or the injury severity of the accidents were applied and evaluated in this research. The KNN algorithm was used to predict the severity of road accidents based on the characteristics of each accident. KNN, chosen for its simplicity and adaptability to medium-sized datasets, classifies points by determining the majority class among their k nearest neighbors. In the context of accident severity prediction, KNN examines accident characteristics, identifies nearest neighbors, and assigns a severity level based on the majority class among these neighbors. The LR and SVM algorithms were also used. These algorithms are known by their robustess, and they were often used for classification and prediction. The NB classifier was also chosen to be used in our research for its speed and efficiency in making predictions with large datasets including independent features. Then the ANN with back propagation was also employed in this research to predict the severity of accidents since it is a very powerful algorithm to provide solutions to complex problems. In our case, the ANN included 15 nodes as input layers, 2 hidden layers with 6 nodes, and output layers with single nodes. The rectified linear unit (ReLU) activation function was applied to introduce non-linearity into the model. Binary cross-entropy loss, suitable for binary classification problems, was used for fatality prediction. For injury severity prediction, which involves multi-class classification, categorical cross-entropy loss was employed. Random forest algorithm was also applied on the dataset to find the most important factors that contributed to increase the accidents severity.

3. RESULTS AND DISCUSSION

The dataset was divided into training (80%) and testing (20%) sets. This division ensures that models will not only memorize the training data used for the training phase but can generalize its predictions to new observations. The model's prediction was assessed on the test set for all the algorithms as shown in Table 3, with accuracy serving as the main performance metric. A high accuracy implies that the model can effectively classify road accident severity on unseen data.

Features were also standardized using the standard deviation method (normalization method for the ANN). This process is crucial to ensure that each feature contributes equitably to the prediction, thus preventing differences in scale from biasing the model in favor of certain variables. Following the training phase, the methods were employed to predict the severity of the accident on the testing data.

3.1. Machine learning models evaluation

The model's effectiveness was assessed using the different metrics, F1-score, recall, precision, and accuracy gauging the ratio of correct predictions to the total predictions. All the mentioned metrics were calculated from the confusion matrix generated by each model using the (1) to (3).

$$Recall = \frac{\mathrm{TP}}{TP+FN} \tag{1}$$

$$Precision = \frac{\mathrm{TP}}{TP+FP} \tag{2}$$

$$F1-score = \frac{2\times Pecision\times Recall}{Pecision+Recall} \tag{3}$$

Where TP is true positive, FP is false positive, TN is true negative, and FN is false negative. These are the values provided by the coefficient matrix for each model. The real accuracy was calculated using the (4).

$$Real\ accuracy = \frac{\mathrm{TP+FN}}{TP+TN+FP+FN} \tag{4}$$

A summary of results for all machine learning models used in this research to predict the fatality or the injuries severity in road accidents on Morocco is shown in Tables 3 and 4. Results showed that the most effective models that gave the best accuracy to predict accidents fatality or accidents injuries severity were the SVM (0.99 accuracy for accident fatality and 0.7 accuracy for accidents injuries severity) and the LR (0.98 accuracy for accident fatality and 0.7 accuracy for accidents injuries severity). ANN gave also a high accuracy for the accident’s fatality (0.98) while it gave the lowest accuracy for the injury’s severity (0.57) this might be due to the majority of categorical features used in the dataset. To overcome the challenge of limited real-time traffic information, this predictive model can lead public security forces towards areas with a significant risk of serious accidents, facilitating proactive intervention and therefore limiting fatal accidents from happening.

Table 3. Evaluation metrics comparison for the predictive methods used for accidents fatality

Machine learning model	Precision	Recall	F1-score	Accuracy
SVM	0.99	0.98	0.98	0.99
KNN	0.99	0.96	0.97	0.99
NB	0.99	0.97	0.97	0.97
LR	0.99	0.98	0.98	0.99
ANN	0.98	0.97	0.97	0.99

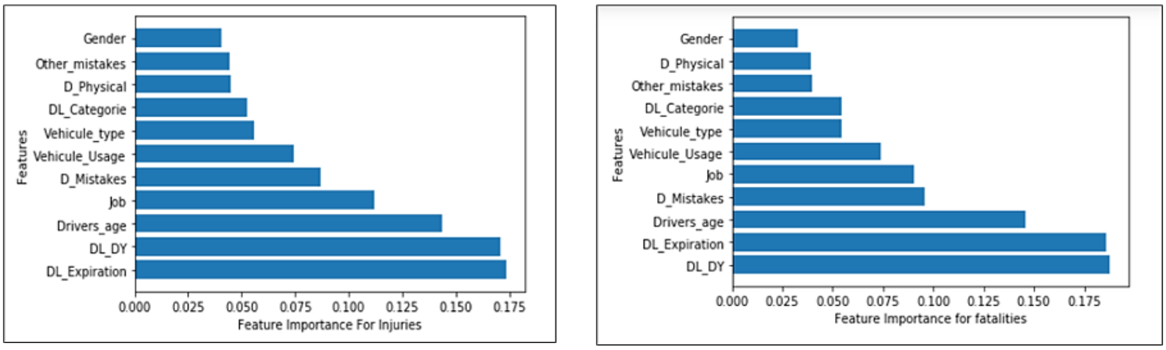
Table 4. Evaluation metrics comparison for the predictive methods used for accidents injuries severity

Machine learning model	Precision	Recall	F1-score	Accuracy
SVM	0.7	0.99	0.82	0.7
KNN	0.71	0.92	0.8	0.68
NB	0.7	0.91	0.8	0.67
LR	0.7	0.99	0.82	0.7
ANN	0.66	0.57	0.62	0.57

Accidents fatality or injuries severity prediction generally searches the relationship between victims' severity and relevant factors (such as human, vehicle, and equipment or environmental factors). A comprehensive analysis of all factors contributing to the severity of accidents is necessary to define the real needs of people in road traffic accidents [25], [26]. This analysis gives critical information to emergency services and traffic managers to implement measures to reduce the side effects of the accident like offering faster medical assistance to people injured in the accidents and therefore, minimizing deaths [5]. To find out the most important factors that can increase road traffic accidents severity from all the given information in our dataset, another machine learning algorithm, random forest, was applied on the same dataset. Results will be discussed in the next section.

3.2. Embedded feature selection in machine learning

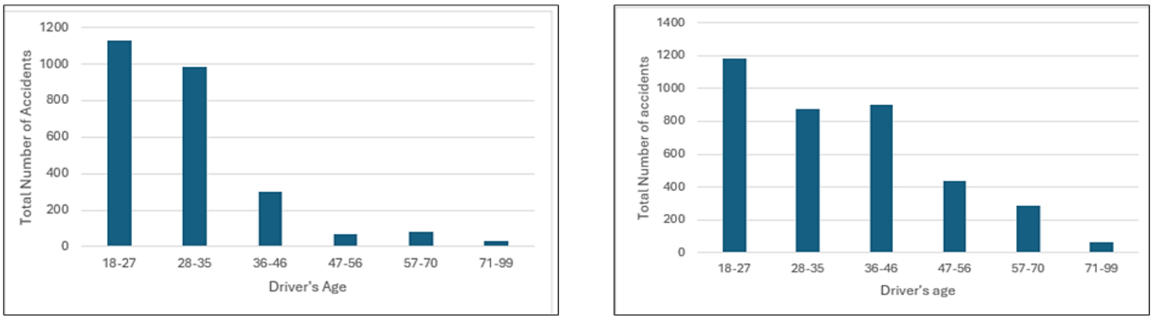
Random forest method collects different results from decision trees and combines them to figure out which attributes are the most important in making a decision. This technique helps understand which features impact the outcome most. Figure 5 show the 10 most important features that leads to the decision selected by the random forest algorithm, where Figure 5(a) shows the fatality prediction and Figure 5(b) shows the injury's severity prediction. Results show that the most important factors that contributes the fatality of accidents or injuries severity are related to the human factor (driver's age, driving license year, driver's profession, and mistakes committed by the driver) followed by the vehicle factor (type of vehicule and vehicule usage). The environment factor (weather, road conditions, and visibility) exhibit a comparatively weak positive correlation with the severity of injuries and fatalities in road traffic accidents.



(a) (b)

Figure 5. Random forest importance feature selection technique applied to the dataset for (a) fatality prediction and (b) injury's severity prediction

One of the most important factor from the human factors was the driver's age. To find out exactly the age range of the driver that leads to increase the severity of the accidents, two columns' charts were deduced from the data set. The chart presented in Figure 6 confirms that the driver's age has an impact on the severity of the accidents, where Figure 6(a) shows the fatal accidents and Figure 6(b) shows the accidents with at least one person suffering serious injuries. The comparison shows that most severe accidents are caused by young drivers (age range between 18 and 27). This can be explained by the fact that young drivers do not have enough experience with driving, road dangers, distances, and estimating speed. They overvalue their driving skills that can lead to speeding and non-adjustment of driving not aware with the road difficulties and surfaces [25].



(a) (b)

Figure 6. Total number by driver's age range of (a) fatal accidents and (b) accidents with at least one person suffering serious injuries

Based on the results found in this research, there should be an adjustment of the driving educational programs and management methodologies in the field of road safety. Giving more importance to the human factor and dividing it into more specific attributes or impact factors will definitely help the traffic authorities to stimulate educational programs especially in the beginner driver education. As an extension to this research, the severity of the accidents can be measured by driver's age and the time of day (daytime or nighttime) to highlight exactly the period of the day when the young and middle-aged drivers are more likely to sustain fatal or severe injury accidents.

4. PROPOSED SOLUTION

SVM is the machine learning method that could predict most accurately the accidents fatality or severity, for this reason, the algorithm was applied again taking into account only the most important factors leading to increase accidents severity. The accuracy of the model was again 99% for accidents fatality and 70% for accidents severity. Nowadays, cars are equipped with ADAS that assist drivers with the safe operation of a vehicle such as automatic emergency braking (AEB) and full auto brake and pedestrian detection (CWAB-PD) systems as stated in [27]. As a proposed solution to this serious real-world problem, a car equipped with a prediction model such as the SVM model, a camera, sensors as well as ADAS systems can perform an early prediction of accidents severity. The car can also be equipped with 3 signals, one indicating that the accident will be fatal, the other two lights will indicate the injuries severity level of the accident that is about to happen.

The SVM model deployed into a car will collect the input from both the driver and the environment. Most of the variables are static and can be entered into the system before driving such as: deriver's age, profession (job), gender, driving license category (DL_category), driving license expiration date (DL_Expiration), vehicle usage, and vehicle type. For the other factors, a camera will be associated with the system to scan the patient's phase. Fatigue detection was also performed using deep learning algorithms [28]. In our research, we proposed that the information is sent to a fuzzy logic algorithm that can provide driver's physical conditions as an output (disability, drunk, narcotics+drugs asleep, tired, sudden sickness or others). Car camera along with sensors can also scan the surrounding environment and collect the data related to the mistakes committed by the driver (failure to obey red lights or stop signs, failure to comply with priorities, crossed mixed continuous line, unreported/defective irregular operation, was driving without precautions, speeding, was driving in a prohibited area, defective stop or parking or others) and mistakes not related to the driver (animals on the road, obstacle abandoned on the roadway, tire enclosure, dangerous hole in the roadway, accidental breakage of the windshield, work site not reported, or others). The data then can be sent to the model and an early prediction of the accident's severity will be determined. The core of our setup involved a simulated vehicle environment equipped with a 3 visual signaling system, one indicating that the accident will be fatal, the other two lights will indicate the injuries severity level of the accident that is about to happen. The design of this setup prioritizes reproducibility, allowing other researchers to replicate our findings and extend this work. The proposed model is summarized in the flowchart shown in Figure 7. It implements a SVM to predict accident severity and outlines the decision-making process at each stage.

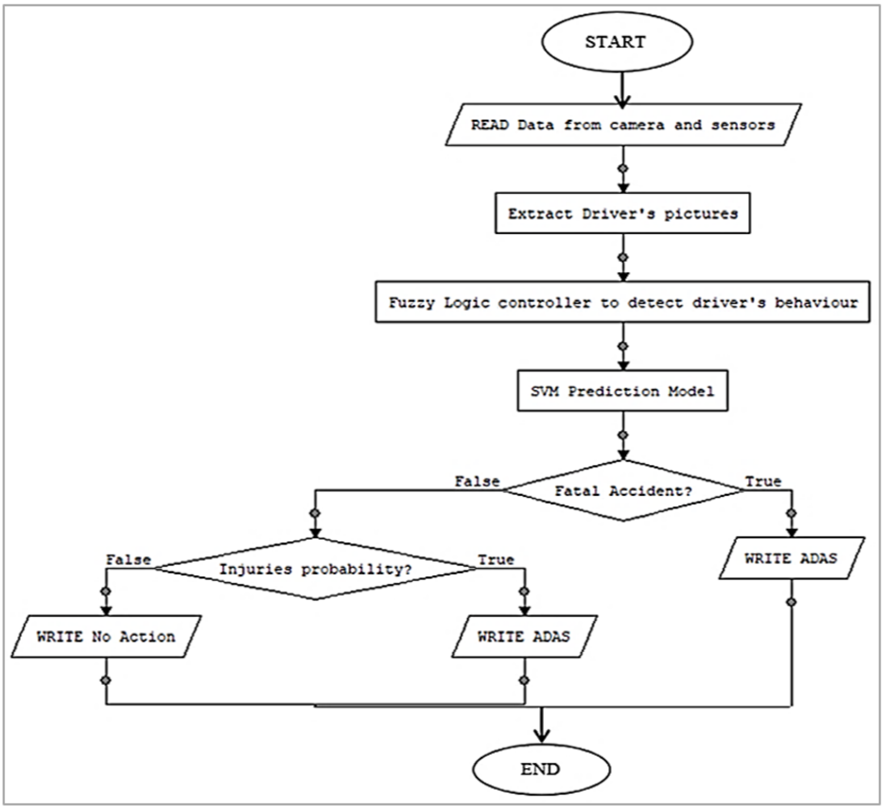


Figure 7. Flowchart of the proposed SVM-based algorithm for accident severity prediction

Another advice to the traffic authorities, there should be also an adjustment of the driving educational programs and management methodologies in the field of road safety. Giving more importance to the human factor and dividing it into more specific attributes or impact factors will definitely help the traffic authorities to stimulate educational programs especially in the beginner driver education. As an extension to this research, the severity of the accidents can be measured by driver's age and the time of day (daytime or nighttime) to highlight exactly the period of the day when the young and middle-aged drivers are more likely to sustain fatal or severe injury accidents. The population growth in Morocco can also be an important factor to consider in data analytics as mentioned [12].

Despite its potential, the application of machine learning for predicting road accidents fatality or injuries severity remains an area with considerable scope for further development. Since the data volumes increase and computing capacity becomes more powerful and affordable, road accidents severity prediction can be performed with complex machine learning models such as deep learning frameworks (deep belief network or recurrent neural network) as used [29]. Deep learning software can analyze big and complex data sets to predict the severity in road accidents faster and more accurately than humans. This ongoing evolution in computational capabilities underscores the importance of reproducible methodologies, allowing other researchers to replicate and further extend these predictive models to enhance road safety outcomes globally.

5. CONCLUSION

This paper uses machine learning algorithms to predict the severity of accidents using a real dataset on accidents in the country of Morocco. The ability to predict and understand crashes has the potential to radically transform crash prevention and reduction efforts, saving lives and reducing the economic and social costs associated with these tragedies. The paper uses a combination of knowledge between artificial intelligence, statistics, and geographic information systems. It uses real data provided by the NARSA for two years. Despite the lack of information on some variables of great interest and unprovided clear coding for other variables, the data could be used, and results were obtained. This paper consists of four fundamental steps: i) data cleaning and preprocessing phase, ii) descriptive statistics to visualize the data and find correlations between the different variables, iii) application of different machine learning algorithms (KNN, LR, SVM, NB, and ANN) to predict the severity of the accidents (fatality or injury's severity) in the country with comparison of the models' performance, and iv) application of the random forest algorithm to find the factors that contribute the most to increase the severity in the accidents. It was found from the results that SVM and LR algorithms were the ones that demonstrated promising abilities to give the best prediction. By applying the random forest importance feature selection technique, it was found that the human factor was the one that increases the fatality of accidents or the severity of injuries. As a proposed solution to this problem, our model can be injected into a car system along with an ADAS system and a fuzzy logic system that indicate's human behaviour based on pictures collected by cameras. This semi-autonomous car an perform an early prediction of accidents severity by lightening an alaeming signal light. Future work can be done to combined algorithms together to develop a robust predictive model, capable of assisting competent authorities in the development of evidence-based road safety policies. The model can then be implemented in a digital application that supports the decision-making process of a road safety system in the country of Morocco to better predict the severity of accidents on the roads. Other machine learning models that can deal with rare cases such as XGBoost and AdaBoost can also be integrated in the model. Compared to machine learning, deep learning algorithms can also be applied to road traffic accidents to reach a better performance due to their capabilities to handle complex, high-dimensional data and extract intricate patterns in road traffic accident analysis and prediction.

ACKNOWLEDGMENTS

We would like to express our gratitude to the National Road Safety Agency (NARSA) in Morocco for providing the comprehensive two-year dataset collected from road cameras and radars. The authors alone are responsible for the analysis and interpretation of the data presented in this article.

FUNDING INFORMATION

No funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Halima Drissi Touzani	✓	✓	✓	✓	✓	✓		✓	✓	✓				
Sanaa Faquir	✓	✓	✓	✓			✓			✓	✓		✓	
Ali Yahyaouy	✓			✓			✓			✓		✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

ETHICAL APPROVAL

The data analyzed in this research relates to aggregated, publicly reported road accidents and fatality statistics provided by NARSA. As this research involved the analysis of pre-existing, non-identifiable, and aggregated human-related data and did not involve the collection of, or interaction with, human subjects or identifiable private information, it was determined to be exempt from formal review. The overall research design and proposal for solutions align with the ethical tenets of the Helsinki Declaration regarding the consideration of human welfare.

DATA AVAILABILITY

The data that support the findings of this study are available on request from the corresponding author, [HDT]. The data, which contain information that could compromise the privacy of research participants, are not publicly available due to certain restrictions.

REFERENCES

[1] CODE delaroute.io, "Road accidents | causes and consequences (in French: *Accident de la route | causes et conséquences*)," *Canopée SAS*. 2023. Accessed: Mar. 10, 2025. [Online]. Available: https://codedelaroute.io/blog/accident-de-la-route/

[2] K. Valášková, T. Kliestik, and M. Mišanková, "The role of fuzzy logic in decision making process," in *2nd International Conference on Management Innovation and Business Innovation (ICMIBI 2014)*, 2014, doi: 10.5729/lnms.vol44.143.

[3] D. Ennajih, M. Elgameh, A. Salik, A. Echchelh, and A. Chaouch, "Study of the influence of parameters of road safety on the road accidents," *International Journal of Research Studies in Science, Engineering and Technology*, vol. 2, no. 4, pp. 22–27, 2015.

[4] C. Yang, J. Jiang, J. Zhou, M. Hitosug, and Z. Wang, "Traffic safety and public health in China – Past knowledge, current status, and future directions," *Accident Analysis & Prevention*, vol. 192, Nov. 2023, doi: 10.1016/j.aap.2023.107272.

[5] P. Infante *et al.*, "Prediction of road traffic accidents on a road in portugal: a multidisciplinary approach using artificial intelligence, statistics, and geographic information systems," *Information*, vol. 14, no. 4, Apr. 2023, doi: 10.3390/info14040238.

[6] H. Kasmi, "Fatal road accidents in Morocco: who is to blame?," *Morocco World News*. 2012. Accessed: Mar. 10, 2025. [Online]. Available: https://www.moroccoworldnews.com/2012/10/149054/fatal-road-accidents-in-morocco-who-to-blame/

[7] ITF, "Road safety country profiles: Morocco 2023," *International Transport Forum*, 2024. [Online]. Available: https://www.itf-oecd.org/sites/default/files/morocco-road-safety.pdf

[8] A. Chand, S. Jayesh, and A. B. Bhasi, "Road traffic accidents: An overview of data sources, analysis techniques and contributing factors," *Materials Today: Proceedings*, vol. 47, pp. 5135–5141, 2021, doi: 10.1016/j.matpr.2021.05.415.

[9] M. I. Sameen, B. Pradhan, H. Z. M. Shafri, and H. B. Hamid, "Applications of deep learning in severity prediction of traffic accidents," in *Global Civil Engineering Conference 2017*, 2019, pp. 793–808, doi: 10.1007/978-981-10-8016-6_58.

[10] 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.

[11] 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.

[12] K. A. R. Torres and J. R. Asor, "Machine learning approach on road accidents analysis in Calabarzon, Philippines: an input to road safety management," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 24, no. 2, pp. 993–1000, Nov. 2021, doi: 10.11591/ijeecs.v24.i2.pp993-1000.

[13] A. P. Akgüngör and E. Doğan, "Estimating road accidents of Turkey based on regression analysis and artificial neural network approach," *Advances in Transportation Studies*, vol. 16, pp. 11–22, 2008.

[14] L.-Y. Chang and H.-W. Wang, "Analysis of traffic injury severity: An application of non-parametric classification tree techniques," *Accident Analysis & Prevention*, vol. 38, no. 5, pp. 1019–1027, Sep. 2006, doi: 10.1016/j.aap.2006.04.009.

[15] M. Guo *et al.*, "A study of freeway crash risk prediction and interpretation based on risky driving behavior and traffic flow data," *Accident Analysis & Prevention*, vol. 160, Sep. 2021, doi: 10.1016/j.aap.2021.106328.

[16] I. C. Obasi and C. Benson, "Evaluating the effectiveness of machine learning techniques in forecasting the severity of traffic accidents," *Heliyon*, vol. 9, no. 8, 2023, doi: 10.1016/j.heliyon.2023.e18812.

[17] A. Chakraborty, D. Mukherjee, and S. Mitra, "Development of pedestrian crash prediction model for a developing country using artificial neural network," *International Journal of Injury Control and Safety Promotion*, vol. 26, no. 3, pp. 283–293, Jul. 2019, doi: 10.1080/17457300.2019.1627463.

[18] M. Jiber, I. Lamouik, Y. Ali, and M. A. Sabri, "Traffic flow prediction using neural network," in *2018 International Conference on Intelligent Systems and Computer Vision (ISCV)*, Apr. 2018, pp. 1–4, doi: 10.1109/ISACV.2018.8354066.

[19] 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.

[20] L. J. -Wook and J. D. Wolfensohn, "Global status report on road traffic injury prevention: summary (in French: *Rapport mondial sur la prévention des traumatismes dus aux accidents de la circulation : résumé*)," *Institutional Repository for Information Sharing*, 2004. [Online]. Available: https://iris.who.int/server/api/core/bitstreams/d35b8fd0-6a90-4e8e-9cba-ea07287fe0dd/content

[21] F. I. Ilunga, "Education routière: une nouvelle approche du comportement de sécurité chez les usagers de la route au Maro," *Arab Journal o Psychology*, vol. 3, no. 2, pp. 130–139, 2018.

[22] H. D. Touzani, S. Bollouz, S. Faquir, and A. Yahyaouy, "Study and analysis of accidents database for the years 2015–2016," in *Digital Technologies and Applications*, Springer, Cham, 2022, pp. 145–152, doi: 10.1007/978-3-031-02447-4_15.

[23] H. D. Touzani, S. Faquir, and A. Yahyaouy, "Data mining techniques to analyze traffic accidents data: Case application in Morocco," in *2020 Fourth International Conference On Intelligent Computing in Data Sciences (ICDS)*, Oct. 2020, pp. 1–4, doi: 10.1109/ICDS50568.2020.9268729.

[24] H. D. Touzani, S. Faquir, S. Senhaji, and A. Yahyaouy, "A proposed solution to road traffic accidents based on fuzzy logic control," in *Digital Technologies and Applications*, Cham, Switzerland: Springer, 2021, pp. 57–63, doi: 10.1007/978-3-030-73882-2_6.

[25] K. Bucsuházy, E. Matuchová, R. Zůvala, P. Moravcová, M. Kostíková, and R. Mikulec, "Human factors contributing to the road traffic accident occurrence," *Transportation Research Procedia*, vol. 45, pp. 555–561, 2020, doi: 10.1016/j.trpro.2020.03.057.

[26] R. L. Khairina, "Overview of latent and active failure factors in Indonesian road transportation accidents in 2019 – 2021," *National Journal of Occupational Health and Safety*, vol. 3, no. 2, 2022, doi: 10.59230/njohs.v3i2.6298.

[27] E. Coelingh, A. Eidehall, and M. Bengtsson, "Collision warning with full auto brake and pedestrian detection - a practical example of automatic emergency braking," in *13th International IEEE Conference on Intelligent Transportation Systems*, Sep. 2010, pp. 155–160, doi: 10.1109/ITSC.2010.5625077.

[28] H. Jia, Z. Xiao, and P. Ji, "Fatigue driving detection based on deep learning and multi-index fusion," *IEEE Access*, vol. 9, pp. 147054–147062, 2021, doi: 10.1109/ACCESS.2021.3123388.

[29] V. Sherimon *et al.*, "An overview of different deep learning techniques used in road accident detection," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 11, 2023, doi: 10.14569/IJACSA.2023.0141144.

BIOGRAPHIES OF AUTHORS

Mrs. Halima Drissi Touzani is a Ph.D. candidate in the Department of Computer Science/LISAC Laboratory at the Sidi Mohamed Ben Abdellah University in Fes in Morocco. She is also a part-time lecturer in the Department of Computer Science teaching undergraduate courses. Her main research interest lies in the area of artificial intelligence, data analytics, and business intelligence. She can be contacted at email: halima.drissitouzaniwalali@usmba.ac.ma.

Dr. Sanaa Faquir received her Ph.D. from Sidi Mohamed Ben Abdellah University in Fes in Morocco. She is currently an Assistant Professor at Vaughn College of Aeronautics and Technology in New York. She worked as an Assistant Professor at the Faculty of Engineering Sciences at the Private University of Fes (UPF) in Morocco for the past years. She was a lecturer at the College of Science and Engineering at the United Arab Emirates University. Additionally, she taught for the Engineering College at Qatar University in Qatar. She was also a lecturer in the Department of Computer Science and Engineering at California State University, Long beach, United States. Her main research interest lies in the area of data analytics, artificial intelligence, renewable energies, and BI. She can be contacted at email: sanaa.faquir@vaughn.edu.

Dr. Ali Yahyaouy is a Professor at the Faculty of Sciences Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, Morocco. He obtained his Ph.D. degree at University of Technology of Belfort-Montbeliard, France and Sidi Mohamed Ben Abdellah University, Morocco in 2010. He is a member of the LISAC Laboratory in Fes. He has been the coordinator of the International Francophone Master "Web intelligence and data science" since 2017. His research activities mainly concern machine learning, multi-agent systems, smart grids, and ITS. He can be contacted at email: ali.yahyaouy@usmba.ac.ma.