International Journal of Computer Engineering and Data Science (ISSN: 2737-8543)
Volume 3– Issue 1, January - March 2023

Does gender affect the identification of road crash occurrences? An Overview and a comparative study

Soukaina EL FEROUALI[1], Zouhair Elamrani Abou Elassad [2], Abdelmounaîm ABDALI[3]

[1,3] CISIEV team, IT department, Faculty of Science and Technology, Cadi Ayyad University
Marrakech, Morocco

[2] SARS Research team, Compute science department, ENSAS, Cadi Ayyad University
Safi, Morocco

E-mail: soukainaelferouali@gmail.com , z.elamrani@uca.ac.ma, a.abdali@uca.ac.ma

Article history

Received Jun 20, 2023
Revised Jun 24, 2023
Accepted Jun 26, 2023
Published Jun 28, 2023

ABSTRACT

Around the world, traffic accidents are regarded as a large and significant cause of injury and death. Nearly 3,700 people are killed and over 1.3 million individuals lose their lives in collisions involving trucks, vehicles, buses, motorcycles, or pedestrians. This article tries to identify the important causes of traffic accidents involving both men and women as well as the methods that have been suggested and put into practice based on the literature study. In order to determine how gender impacts the frequency of traffic accidents, a survey and a comparative study were conducted in this work. According to the findings, the factors that have been studied for accident causes in urban areas include speed, age, and gender. On rural roads, speed has been recognized as the primary cause of collisions, particularly among men, while age and lack of experience have been noted as influencing factors in women's traffic incidents. Because machine learning models are effective at predicting crashes, they have been utilized in the majority of research.

Keywords: ***Road safety, Accident severity, Risk factors, Machine learning.***

I. INTRODUCTION

Road safety is an essential subject for scientific research, and one of these major aspects with a high degree of importance is road accidents, which are a very useful area of study. Road crashes are a major public health problem, so it is very important to understand the factors responsible for them and the consequences that follow to make the correct decisions to reduce the severity of these accidents. The latter are responsible for 1.35 million deaths in the world each year, an average of 3,287 deaths per day, to which must be added a very large number of injuries representing 20 to 50 million disabled people [2] [19], which explains the increase in social and economic problems. In 2018, more than 1.5 million road users died and more than 50 million others were injured or permanently disabled, particularly among vulnerable road users (i.e., pedestrians and motorcyclists) [20]. Traffic accidents are additionally classed as the ninth greatest cause of mortality worldwide, according to the World Health Organization (WHO) [20]. Based on present trends, it is predicted that by 2030, there would be about two million road fatalities annually. Human factors are the main cause of more than 90 % of road accidents, according to a number of research studies [19]. The potentialis therefore very important in terms of prevention. However, these factors are difficult to identify, which is why the focusso far has been on vehicle and infrastructure improvements. Traffic accident prediction models are very powerful tools thatare widely used to determine the factors associated with and responsible for these accidents [9]. The purpose of this study is to present the effect of various causal factors on crash severity for both male and female drivers, and the machine learning and deep learning techniques that have been proposed and applied to reduce the severity rate. The reason for focusing on crash severity rather than the number and type of crashes is due to the importance of understanding the factors leading to increased crash severity and the respective injuries sustained as a result. These factors differ for men and women, which makes the study all the more necessary. The article would present thecontinuity of a study carried out previously in an article that was based on the critical factors responsible for road accidents,of which four main categories were put forward for the study, namely environmental, technical, demographic, and habitual factors. Each category of these factors is related to its aspects as shown in Figure 2, so this study will have a direct impact on these critical factors mentioned previously with the aim of making a comparison

between male and female drivers in their behaviors on the road. The document is organized as follows: The second part will present the related research on traffic accidents, which will be divided into two sections: one on the causes of traffic accidents and another on the state of the art, which includes the methods employed and the findings of the numerous studies conducted, and finally a conclusion and recommendations for future research.

II. RELATED WORK

In this section, several studies will be presented in order to compare the factors responsible for road accidents in the two categories of drivers based on research results. Two figures 1 and 2 above highlight first the essential features of worldwide road safety and then the factors that contribute to crashes particularly.

A traffic crash is the result of a breakdown in the balance of human, environmental, and technical elements, which represent the three components of road safety. These elements are shown in the figure above.

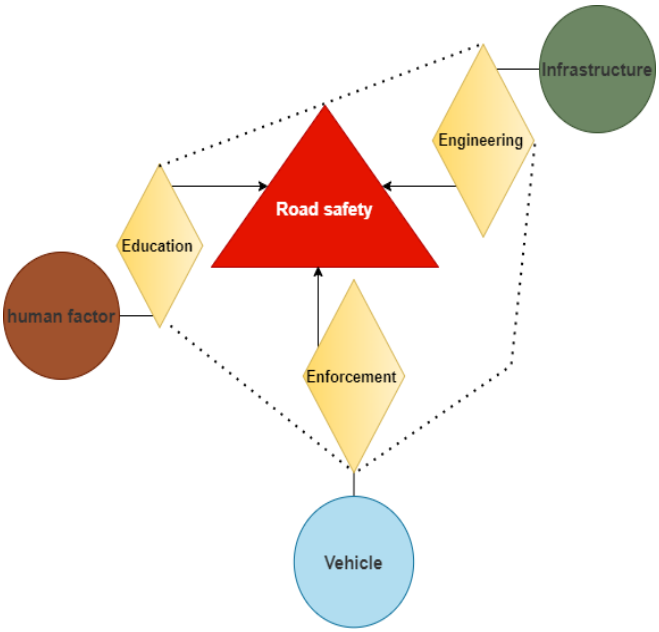

Figure 1:The three components of road safety

Each of these components refers to one of three different variables that contribute to the promotion and maintenance of road safety which are: engineering, enforcement, and education components. The engineering element is concerned with the design and maintenance of road infrastructure, vehicles, and related systems in order to improve safety. Road design, traffic flow management, road signs, and transportation planning are all part of it. Engineering initiatives aim to lower the probability of accidents, lessen the severity of accidents, and promote safer conditions for all road users. The enforcement component is responsible for establishing and executing traffic laws and regulations. It entails the operations of law enforcement agencies to monitor and enforce compliance with road safety laws. This includes measures such as speed limit enforcement, drunk driving detection, seat belt and helmet requirements, and verifying compliance with vehicle rules. Effective enforcement discourages risky behavior and assures accountability for violators. The education component strives to raise road users' awareness, knowledge, and abilities. It includes teaching the public about road safety standards, proper driving behavior, and the risks involved with various means of transportation. Education campaigns, school programs, and driver training all play an important part in encouraging responsible driving and enhancing understanding of road safety rules.

A. *Categories of critical factors*

Starting with the environmental factor [2] [5] [6] [11] [12] and its different aspects that can influence the driver's behavior like weather conditions and pollution, which causes fatal accidents. The demographic characteristics [1] [3] [5] [7] [11] [14] of the drivers were treated as follows: health, age, gender,and education. Then, the habit factor [2] [3] [4] [7] [11] andits aspects, including fatigue, alcohol consumption, helmet, and seat-belt use, and cell phone use as the main factors influencing driver behavior, and finally, the technical factors [2] [5] [6] [11], which represent one of the most studied and used categories in several studies, are due to their importance and their direct influence on the driver either by the state ofhis vehicle, which can influence negatively while driving, or the infrastructure, which is considered an external factor.

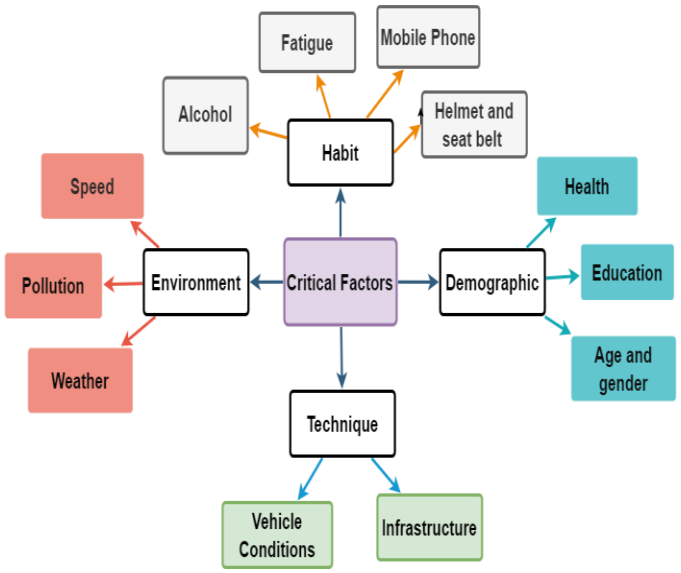

Figure 2: Critical factors

B. *State of the art*

Several studies have been able to determine and analyze the factors responsible for road accidents, among them, that of Massie et al [17], who concluded that serious and fatal accidents are more likely for male drivers than for female drivers [16]. As revealed in one study, two important aspects of road safety, namely the prediction of injury severity and the identification of influential parameters (significant risk parameters) have been the subject of previous research [18].

Four studies were conducted in order to see the parameters influencing each category, namely those of the driver, the road, the environment, and the vehicle, and found that (age, alcohol consumption, and seat-belt use) are the parameters influencing the drivers, (the speed limit) influences the road, (weather conditions,

lighting, and time of the accident) influence the environment, and finally (the type of vehicle and its condition) influence the vehicle, respectively [18]. Another study was revealed to examine the factors affecting the severity of urban accidents in male and female drivers and found that inexperience, loss of control, and mechanical defects were the main causes of road accidents in women, whereas speed and left turns were the main risk factors in men [16]. According to the literature review [22], the authors found several studies on factors affecting crash severity that adopted different types of mathematical models that are grouped into different categories, including road, external environment, and driver. Different accidents are caused by men in principle, who are between the ages of 18 and 44 [22]. These studies have approved that age is a factor affecting road accidents in the first degree. Analyzing and predicting accidents requires the use of machine learning and deep learning algorithms. However, statistical analyses of accidents were also included in the studies.

a) Interpretation of the findings from several studies reported in the state of the art that make use of machine learning algorithms:

The C4.5 algorithm is a supervised classification algorithm with significant processing power. It presents simple classification rules that can be interpreted by humans and serves as a decision tree classifier that may be used to produce a decision based on a specific sample of data (univariate or multivariate predictors). Random forests, or random decision forests, are a learning method for classification and regression that works by building a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by the most trees. For regression tasks, the average or mean prediction of individual trees is returned. The prediction accuracy of the RF is computed by aggregating the results of the DTs, which corresponds to the majority vote for classification or the mean for regression. In machine learning, instance-based learning (sometimes called memory-based learning) is a family of learning algorithms that, instead of performing explicit generalization, compare new instances of problem with instances seen in formation, which were stored in memory.

Different studies have used other machine learning algorithms to predict the severity of accidents and identify impact factors such as: logistic regression represents a statistical model to study the relationships between a set of qualitative variables Xi and a qualitative variable Y. A logistic regression model also predicts the probability of an event occurring (value of 1) or not (value of 0) from the optimization of the regression coefficients. This result always varies between 0 and 1.

Table I displays the findings from a study [16] that investigates the variables influencing both male and female drivers' severity of urban crashes. Observations were split into two groups according to the gender of the drivers. Using the binary Logit model, two crash severity models were calculated for each group, with the dependent variables being damage-only crashes and injury/fatal crashes.

The findings indicated that factors like age, less-educated drivers, weekends, nights, and intersections have increased crash severity in both male and female drivers. Men were more severely affected by breaking the law, speeding, and veering to the left, while women were more severely affected by inexperience, loss of control, and mechanical problems with the vehicle.

Table I: FACTORS AFFECTING ROAD ACCIDENTS AMONG WOMEN AND MEN

	Gender Categories	
Critical Factors	Male	Female
Increasing Age	✓	✓
Untrained drivers	✓	✓
Week-end	✓	✓
Night	✓	✓
Intersections	✓	✓
The violation of the rules	✓	
Speed	✓	
Diversion to the left	✓	
Lack of experience		✓
Loss of control		✓
Mechanical defects		✓

Table II shows the results of three machine learning methods applied in a study [18] which has the aim of analyzing the severity of injuries suffered by drivers of different age and gender groups, and which shows that the category of older men(OM) is the most exposed to road accidents with an accuracy of 80.49 % using the C4.5 algorithm, 84.85 % using the Instance-based learning (IB) and 85.37 % using the Random Forest (RF) algorithm; therefore, compared to the other algorithms used, the Random Forest approach is accurate in terms of precision.

The findings shown in Table III illustrate the results of a research study [22] carried out using the logistic regression model on the variable representing the severity of the injury. A binary logistic regression was carried out, and three models were created. The accident severity variable was recoded into two levels: 0 for accidents with fewer than one injury (low severity accidents), and 1 for incidents with more injuries or fatalities (severe and fatal accidents). The model is most successful in predicting the type of accident, and the overall percentage, which is commonly regarded to be about 70 %, continues to be high and satisfactory. By applying the first model, where the injuries are not very serious, the second model

where the injuries are somewhat serious, or the third model, where the injuries are extremely serious, we can infer that gender is not a significant variable; nevertheless, age is a significant variable by 10 % in very serious injuries.

Table II: CONFUSION MATRICES OF C4.5, IB, AND RF CLASSIFIERS FOR FOUR GROUPS OF DRIVERS IN PERCENT

Gender and age categories	Models	Accuracy
Older Female (OF)	C4.5 IB RF	73.30 83.70 84.39
Older Male (OM)	C4.5 IB RF	80.49 84.85 85.37
Younger Female (YF)	C4.5 IB RF	73.55 79.55 79.96
Younger Male (YM)	C4.5 IB RF	75.58 83.26 83.68

Table III: RESULTS OF LOGISTIC REGRESSION MODELS (***SIGNIFICANCE AT 1%LEVEL, **SIGNIFICANCE AT 5%LEVEL, *SIGNIFICANCE AT 10%LEVEL, N.S(NOT SIGNIFICANT)

Variables	Model 1	Model 2 (severity= 0)	Model 3 (severity=1)
Location	***	***	*
Street	***	***	***
Weather condition	**	**	n.s
Day of the week	***	***	***
Hour	***	***	***
Season	***	***	***
Age	***	***	*
Gender	n.s	n.s	n.s
Vehicle	***	***	**

b) Interpretation of the results from various deep learning algorithms research:

Deep learning is one of the main technologies of machine learning that is part of artificial intelligence. With Deep Learning, we are talking about algorithms capable of mimicking the actions of the human brain through artificial neural networks. These networks are composed of tens to hundreds of "layers" of neurons, each receiving and interpreting information from the previous layer. Deep learning models tend to work well with large amounts of data, whereas more traditional machine learning models stop improving after a saturation point, which makes a difference between the two methods.

Studies have proposed several deep learning approaches, among them, a study [23] that used a comprehensive analytical framework based on a deep learning model called "stacked sparse autoencoder" (SSAE) to predict the severity of injuries caused by traffic accidents based on the contributing factors. The key idea of this method is to integrate various analyses by performing processing and analysis of the corresponding data by different machine learning approaches such as Catboost which is used to analyze the importance and dependence of the factors contributing to the severity of injuries and remove the factors of weak correlation, the approach (k-means clustering) was used to classify the data into several classes and finally and with high correlation factors, the SSAE-based deep learning model was built to perform injury severity prediction in each data class[23].

TABLE IV lists the statistical findings (i.e., frequencies and percentages) for some driver characteristics, such as age and gender, and how often they occur in traffic accidents resulting in serious and moderate injuries.

Table IV: DESCRIPTIVE STATISTICS FOR AGE AND GENDER FEATURES

		Serious/ Fatal Injury Severity		Slight Injury Severity	
Driver characteristics	Variables	Freq	%	Freq	%
Gender	Male	79,960	71,53	416,383	63,21
	Female	31,829	28,47	242,389	36,79
Age	16-20	9880	8,84	55,944	8,49
	21-25	13,790	12,34	85,034	12,91
	26-35	22,603	20,22	150,334	22,82
	36-45	20,497	18,34	132,469	20,11
	46-55	19,521	17,46	113,074	17,16
	56-65	12,562	11,24	65,271	9,91
	66-75	7288	6,52	34,692	5,27
	Over 75	5648	5,05	21,954	3,33

III. DISCUSSION OF LITERATURE

The results of this study [16] show that male and female drivers have different characteristics that influence crash severity, which may be because of the ways in which they drive. Authorities and safety experts are cautioned by this problem to evaluate these variations before taking any actions or formulating any policies to enhance safety. Therefore, it may be suggested that future research on this topic investigate driving and behavioral variations between

males and females on suburban roads and compare them to those on urban streets. Studying the potential differences in driving behaviors at various urban areas (such as crossroads and motorways, for example) might also be a crucial area of research. As a last suggestion, a future study should compare how the gender of the driver influences crash severity in industrialized and developing nations.

Interventions on elements relating to the road seem to be necessary regardless of the seriousness of an accident to increase road safety [22]. Therefore, by enhancing road crossings, designing urban roads better, and encouraging the use of one-way roads, traffic safety might be raised. Traffic conditions must be improved in order to increase road safety, particularly at night, and minimize the number of major accidents. Lastly, driver-related issues should also be better managed in order to decrease the frequency of serious accidents.

IV. CONCLUSION AND PERSPECTIVE

In order to make informed decisions that will lessen the severity of traffic accidents, it is essential to understand the causes and consequences of these events. Based on the various criteria described in the literature review, we may conclude that male drivers are more likely to be involved in major and fatal crashes than female drivers.

Future research will concentrate on additional elements that cause traffic accidents, including the type of vehicle, the type of road, and driver distraction factors. These parameters will be examined once more to determine their impact on crash occurrences.

REFERENCES

[1] S. Regev, J. J. Rolison, and S. Moutari, “Crash risk by driver age, gender, and time of day using a new exposure methodology,” J. Safety Res., vol. 66, pp. 131–140, 2018.

[2] M. Mohanty and A. Gupta, “Factors affecting road crash modeling,” J. Transp. Lit., vol. 9, no. 2, pp. 15–19, 2015.

[3] O. S. Liang and C. C. Yang, “How are different sources of distraction associated with at-fault crashes among drivers of different age gender groups?,” Accid. Anal. Prev., vol. 165, no. February 2021, p. 106505, 2022.

[4] L. T. Truong, H. T. T. Nguyen, and R. Tay, “A random parameter logistic model of fatigue-related motorcycle crash involvement in Hanoi, Vietnam,” Accid. Anal. Prev., vol. 144, no. June, p. 105627, 2020.

[5] X. Zhai, H. Huang, N. N. Sze, Z. Song, and K. K. Hon, “Diagnostic analysis of the effects of weather condition on pedestrian crash severity,” Accid. Anal. Prev., vol. 122, no. November 2018, pp. 318–324, 2019.

[6] Z. Ma, S. I. J. Chien, C. Dong, D. Hu, and T. Xu, “Exploring factors af- fecting injury severity of crashes in freeway tunnels,” Tunn. Undergr. Sp. Technol., vol. 59, pp. 100–104, 2016.

[7] J. J. Rolison, S. Regev, S. Moutari, and A. Feeney, “What are the factors that contribute to road accidents? An assessment of law en- forcement views, ordinary drivers’ opinions, and road accident records,” Accid. Anal. Prev., vol. 115, no. August 2017, pp. 11–24, 2018.

[8] P. B. Silva, M. Andrade, and S. Ferreira, “Machine learning applied to road safety modeling: A systematic literature review,” J. Traffic Transp. Eng. (English Ed., vol. 7, no. 6, pp. 775–790, 2020.

[9] N. Mor, H. Sood, and T. Goyal, “Application of machine learning technique for prediction of road accidents in Haryana-A novel ap- proach,” J. Intell. Fuzzy Syst., vol. 38, no. 5, pp. 6627–6636, 2020.

[10] S. S. Yassin and Pooja, “Road accident prediction and model interpre- tation using a hybrid K-means and random forest algorithm approach,” SN Appl. Sci., vol. 2, no. 9, pp. 1–13, 2020.

[11] A. Mohammed et al., “A Landscape of Research on Bus Driver Be- havior: Taxonomy, Open Challenges, Motivations, Recommendations, Limitations, and Pathways Solution in Future,” IEEE Access, vol. 9, pp. 139896–139927, 2021.

[12] A. Abdulhafedh, “Road Crash Prediction Models: Different Statistical Modeling Approaches,” J. Transp. Technol., vol. 07, no. 02, pp. 190–205, 2017.

[13] Z. Elamrani Abou Elassad and H. Mousannif, Understanding driving behavior: Measurement, modeling and analysis, vol. 915. Springer International Publishing, 2019.

[14] C. Lin, D. Wu, H. Liu, X. Xia, and N. Bhattarai, “Factor identification and prediction for teen driver crash severity using machine learning: A case study,” Appl. Sci., vol. 10, no. 5, 2020.

[15] S. Ul Hassan, J. Chen, A. A. Shah, and T. Mahmood, “Acci- dent detection and disaster response framework utilizing IoT,” Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 3, pp. 379–385, 2020.

[16] H. Razi-Ardakani, A. Ariannezhad, and M. Kermanshah, “A study of sex differences on road crash severity,” Int. Conf. Civil, Struct. Transp. Eng., no. 113, pp. 113.1-113.14, 2018.

[17] D. L. Massie, P. E. Green, and K. L. Campbell, “Crash involvement rates by driver gender and the role of average annual mileage,” Accid. Anal. Prev., vol. 29, no. 5, pp. 675–685, 1997.

[18] S. Mafi, Y. AbdelRazig, and R. Doczy, “Machine Learning Methods to Analyze Injury Severity of Drivers from Different Age and Gender Groups,” Transp. Res. Rec., vol. 2672, no. 38, pp. 171–183, Dec. 2018.

[19] C. Charbit, “Les facteurs humains dans les accidents de la circulation: un potentiel important pour des actions de pre´vention,” En Ligne], Www. Fond. Maif. Fr/ . . . , 1997, [Online].

[20] “World Health Organization,WHO, (2015).”

[21] M. A. Rahim and H. M. Hassan, “A deep learning based traffic crash severity prediction framework,” Accid. Anal. Prev., vol. 154, no. February, p. 106090, 2021.

[22] L. Eboli, C. Forciniti, and G. Mazzulla, “Factors influencing accident severity: An analysis by road accident type,” Transp. Res. Procedia, vol. 47, pp. 449–456, 2020.

[23] Z. Ma, G. Mei, and S. Cuomo, “An analytic framework using deep learning for prediction of traffic accident injury severity based on contributing factors,” Accid. Anal. Prev., vol. 160, no. July, p. 106322, 2021.