Eco Cities (2023) Volume 4 Issue 1
doi: 10.54517/ec.v4i1.2250

ORIGINAL RESEARCH ARTICLE

Exploring traffic accidents patterns: Spatial distribution and socio-economic determinants

Pires Abdullah[1,2,*], Tibor Sipos[1,3]

[1] *Department of Transport Technology and Economics, Faculty of Transportation Engineering and Vehicle Engineering, Budapest University of Technology and Economics, H-1111 Budapest, Hungary*

[2] *Department of Spatial Planning, College of Spatial Planning, University of Duhok, 1006 AJ Duhok, Iraq*

[3] *KTI—Institute for Transport Sciences, Directorate for Strategic Research and Development, 1119 Budapest, Hungary*

*** Corresponding author:** Pires Abdullah, pires.abdullah@edu.bme.hu

ABSTRACT

This study employs a machine learning methodology, specifically the decision tree algorithm, in conjunction with the Quantum Geographic Information System (QGIS), to conduct a rigorous analysis of traffic accident data. The research aims to investigate various factors associated with traffic accidents, with a particular emphasis on their spatial distribution and the socio-economic determinants contributing to recurring accidents caused by drivers. The study focuses on the city of Duhok, located in the Kurdistan Region of Iraq, and utilizes a questionnaire to collect data from drivers regarding accident locations and the frequency of accidents within the past decade (from 2010 to 2020). The findings of the study reveal that the city center experiences the highest concentration of accidents, while severe collisions tend to occur in specific "black spots" scattered across the city's road network. The decision tree model, employed to classify drivers with multiple accidents, identifies the primary causes of accidents as traffic conditions, traffic law violations, and overspeeding. Furthermore, the accident locations are found to be influenced by various factors, including different types of road hierarchy. The age and gender of drivers also contribute to accident patterns. These research findings have practical implications for enhancing road safety measures and reducing the frequency of traffic accidents. The utilization of machine learning techniques, combined with the analysis of spatial data through QGIS, provides a comprehensive understanding of the underlying factors contributing to accidents. Moreover, this research contributes novel insights to the field of road traffic accidents and safety, particularly in the context of the city of Duhok in Kurdistan Region, Iraq, and provides a valuable reference for future studies in the domain of road safety and urban planning.

Keywords: road accidents; spatial analysis; socioeconomic factors; decision tree classifier

1. Introduction

Effective management of transportation and traffic systems in modern cities presents a complex and ongoing challenge, necessitating comprehensive approaches to identify and address key issues[1]. Among the critical concerns in urban areas, traffic accidents stand out as a significant problem, stemming from a multitude of interacting factors. While numerous studies have examined various human behaviors that contribute to accident causation, there remains a need for further investigation into specific driver groups and behaviors that

ARTICLE INFO
Received: 18 July 2023 | Accepted: 16 August 2023 | Available online: 22 September 2023

CITATION
Abdullah P, Sipos T. Exploring traffic accidents patterns: Spatial distribution and socio-economic determinants. *Eco Cities* 2023; 4(1): 2250. doi: 10.54517/ec.v4i1.2250

play a crucial role in traffic accidents[2]. Understanding the underlying causes and spatial patterns of accidents is paramount for developing targeted interventions and improving road safety measures.

Existing research has highlighted the importance of speed-related indicators in assessing accident severity, gaining consensus among scholars[3]. However, the incorporation of the severity dimension into modeling crash frequency has received relatively less attention[4]. Notably, there is a stronger correlation between high-risk traffic accidents and factors related to road geometry and traffic conditions[5]. Insufficient traffic conditions and inadequate road infrastructure have been linked to approximately 20%–25% of accidents, as statistically demonstrated[5].

A research study conducted in Tunisia has provided significant insights into collision patterns within the context of a geographical information system (GIS) environment during different time periods[6]. Additionally, another research investigation addresses methodological challenges in the systematic identification of hot zones and probable hot zones for road crashes across an entire road network, while comparing different mapping approaches for practical application[7]. The Kernel Density Estimation (KDE) method has been effectively utilized to detect concentrated areas of road accidents during peak time periods, predict hourly trends in pedestrian-vehicle collisions, and investigate temporal fluctuations in high-risk locations[8]. Moreover, a separate study introduces the innovative concept of “hazardous probable lengths,” aimed at predicting future traffic crashes and enhancing the ability to examine specific lanes for potential hazards[9]. Furthermore, additional studies employ the Spatial Data Analysis Method, specifically Moran’s I and Getis-Ord, to effectively identify spatial concentration locations of “black zones” associated with road accidents and further classifying them based on their respective danger levels[10,11].

Efforts to enhance traffic safety and reduce accidents are evident in various countries. In Turkey, a study utilizes GIS and statistical methods to identify traffic accident hotspots on state roads, pinpointing black spots and reducing accidents[12]. Additionally, studies focusing on different road types have revealed varying rates of injury and death resulting from traffic collisions, with administrative roads having the highest death rate, followed by functional roads, while urban general roads exhibit higher accident rates compared to urban expressways[13,14].

To address the complexities of traffic accidents, it is essential to study behavioral aspects and contributing factors. The decision tree approach provides a valuable methodology for identifying influential factors and constructing predictive models[15]. Through the iterative construction of a decision tree based on survey data, researchers can identify influential variables that impact drivers’ comprehension and decision-making related to the traffic system[15]. This approach helps achieve homogeneous groups concerning the dependent variable by iteratively reducing entropy and achieving consistency with the dataset[16–18]. While previous studies have primarily focused on external factors, such as crash probabilities, driver characteristics, pavement surface conditions, and identifying accident causes[12], there is a noticeable scarcity of studies that explore spatial and causal patterns of collisions, particularly with regards to incorporating socioeconomic factors.

Despite existing research on the adverse effects of traffic accidents, limited attention has been given to exploring the factors contributing to frequent accidents and the influence of socioeconomic factors. This study aims to bridge this gap by conducting a spatial analysis of traffic accidents on the transport network, with a specific focus on identifying socioeconomic factors associated with the occurrence of multiple crashes reported by drivers. By gaining insights into the spatial patterns and factors influencing traffic accidents, this study seeks to inform targeted interventions and improve road safety measures in the urban context. By extending the existing research, this study contributes to a more comprehensive understanding of the complexities surrounding traffic accidents and facilitates evidence-based decision-making to enhance road safety in urban areas.

2. Methodology

The methodology section of this study employed a mixed approach, combining data analysis using Python programming and the decision tree algorithm with spatial analysis using Quantum GIS software. Data was collected through questionnaires administered to drivers in the city of Duhok, Kurdistan Region of Iraq, in 2021, gathering information on traffic accidents, socio-economic characteristics, and accident locations. **Figure 1** illustrates the sequential steps of the study's methodology.

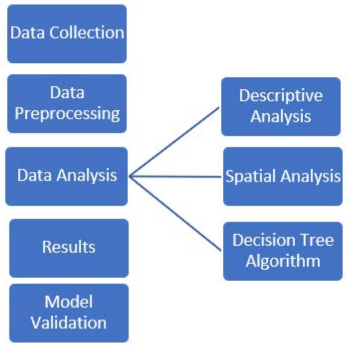

Figure 1. Summary of the methodology steps.

2.1. Study area

The research was conducted in Duhok, a city located in the Kurdistan region of Iraq. The study aimed to cover various districts within the city, encompassing different socio-economic backgrounds and road network characteristics.

2.2. Data collection

A questionnaire-based survey was administered to collect data on traffic accidents. The questionnaire was distributed among diverse groups of residents in the city, ensuring representation from different demographic segments. The questionnaire included questions about driving experiences, behaviors, social and economic impacts, opinions, and general information. Additionally, a city map was provided to assist respondents in identifying the locations of traffic accidents they may have witnessed over the previous 10 years. The questionnaire also gathered information on accident severity, categorizing accidents as severe (involving injuries or fatalities) or non-severe (involving only vehicle or property damage).

2.3. Dataset

The dataset used in the study consisted of 368 samples obtained from the completed questionnaires. The dataset encompassed a range of socio-economic characteristics of the drivers as independent variables. The dependent variable was binary, indicating whether the driver had experienced a single accident (represented as "0") or multiple accidents (represented as "1").

2.4. Data analysis

The Python programming language was utilized for data analysis. A web-based application process was developed using Python, which facilitated the creation of clear and informative reports through effective visualization tools. The decision tree algorithm was implemented using Python programming to classify the drivers based on their accident history and socio-economic characteristics. The decision tree algorithm is a supervised classification technique that utilizes predictive attributes or features to anticipate the class variable[19–21].

2.5. Spatial analysis

Data visualization and representation using computer software offer valuable insights into traffic accident data, enabling comprehensive analysis and supporting decision-making processes. Geographic Information System (GIS) tools, such as shapefiles, allow for the geographical representation of urban road networks and the spatial events occurring on them. Through GIS-based visualization, researchers can identify potential crash risks on road segments, estimate the probability of accidents, and analyze the relationships between various factors[22]. Visual representations, including maps, scatterplots, and heat maps, facilitate the exploration of spatial patterns and the associations between accidents and related variables, such as fatalities or injuries[23].

To identify high-risk areas, or "black spots," for traffic accidents, the Kernel Density Estimation algorithm was applied in QGIS. The algorithm created a heat map using latitude and longitude coordinates obtained from the questionnaire responses. This spatial analysis technique helped visualize the concentration of accidents across the city's road network.

The frequency of accidents and the corresponding fatalities and injuries that occurred in the city during the previous decade have been documented in **Table 1** of a previously published study[24]. The table provides data on traffic accidents, including the number of accidents, the number of people who died as a result of these accidents, and the number of people injured, for each year from 2010 to 2019.

Table 1. Number of accidents and causalities by vehicles in Duhok city[24].

Year	No. of accidents	No. of people dead	No. of injured
2010	362	88	754
2011	707	83	691
2012	629	129	722
2013	1319	260	4290
2014	1132	189	4213
2015	1220	172	3967
2016	1177	205	4003
2017	1152	145	1658
2018	1091	117	1071
2019	1002	98	1102

2.6. The decision tree method

The output generated by the Decision Tree (DT) method illustrates the hierarchical representation of influential factors that affect the occurrence of multiple crashes among drivers. The DT algorithm facilitated the hierarchical representation of these influential factors, highlighting their relative importance. **Table 2** presents the variables and their respective descriptions used in the analysis:

- Age: This variable categorizes individuals into different age groups. The values assigned are 0 for the age group 18–25, 1 for 26–39, 2 for 40–64, and 3 for 65 or older.
- Gender: This variable distinguishes between male (0) and female (1) drivers.
- Education level: This variable represents the educational background of the drivers. The values assigned are 0 for not educated, 1 for secondary school, 2 for university students, and 3 for graduated individuals.
- Location of the crash: This variable identifies the location of the crash within the city.

- Causes: This variable indicates the causes of the traffic accidents. The values range from 1 to 6, where 1 represents traffic conditions and road geometry, 2 represents traffic law violations, 3 represents overspeeding, 4 represents vehicle malfunctions, 5 represents the driver's attitude (recklessness, distracting activities, etc.), and 6 represents other factors such as climate conditions and road facilities.
- Target variable (multi accidents): This variable indicates whether a driver has experienced multiple accidents (1) or only one accident (0).

These variables were analyzed to investigate the relationship between different factors (age, gender, education level, crash location, and causes) and the occurrence of multiple accidents.

Table 2. The dataset names and their numerical description.

Variables	Description
Age	Age groups (0 = 18–25; 1 = 26–39; 2 = 40–64; 3 = 65 or older)
Gender	Male = 0; Female = 1
Education level	(0 = Not educated; 1 = Secondary School; 2 = University Student; 3 = Graduated)
Location of the crash	0 = CBD; 1 = Highway; 2 = Arterials; 3 = Collector; 4 = Local;
Causes	1 = Traffic condition and road geometry; 2 = Traffic low violation; 3 = Over speeding; 4 = Vehicle mal functioning; 5 = Driver's attitude (Recklessness, distracting activities, etc.); 6 = Others (climate condition, road facilities, etc.)
Target variable	-
Multi accidents	0 = one accident; 1 = two or more accidents

2.7. Descriptive analysis

Table 3 presents descriptive statistics for various variables related to traffic accidents based on a dataset of 368 observations. These descriptive statistics provide an overview of the characteristics of the respondents and the nature of traffic accidents in the dataset. The variables included in the analysis are as follows:

Table 3. The dataset names and their numerical description.

	Age	Gender	Owning a car	Education level	Had a road accident	No. of times	Location of the crash	Severity	Causes
Count	368	368	368	368	368	368	368	368	368
Mean	1.173	0.22	0.67	1.81	1.0	1.77	1.82	0.26	3.17
std	1.02	0.42	0.46	1.20	0.0	1.16	1.42	0.43	1.57
min	0.00	0.00	0.00	0.00	1.0	1.00	0.00	0.00	1.00
25%	0.00	0.00	0.00	1.00	1.0	1.00	1.00	0.00	2.00
50%	1.00	0.00	1.00	2.000	1.0	1.00	1.00	0.00	3.00
75%	2.00	0.00	1.00	3.00	1.0	2.00	3.00	1.00	5.00
max	3.00	1.00	1.00	3.00	1.0	6.00	4.00	1.00	6.00

- Age: The average age of the respondents is approximately 1.17, with a standard deviation of 1.03. The minimum age is 0, while the maximum age is 3. The 25th percentile indicates that 25% of the respondents have an age of 0, and the median (50th percentile) age is 1.
- Gender: The variable is binary, with 0 representing males and 1 representing females. The mean value suggests that approximately 23% of the respondents are female.
- Owning a car: This binary variable indicates whether the respondent owns a car or not. The mean value of 0.67 implies that around 67% of the respondents own a car.

- Education level: The respondents' education level is measured on a scale from 0 to 3, representing different levels of education. The mean of 1.82 indicates an average education level of 1. The standard deviation is 1.21, suggesting some variability in education levels among the respondents.
- Had a road accident: This variable is binary, with a constant value of 1 for all observations, indicating that all respondents had experienced a road accident. The Python code implemented in this study aimed to extract a subset of data focusing on drivers who had been involved in traffic accidents. By utilizing the code, the researchers selectively chose the relevant observations from the entire dataset to narrow their analysis to the target population. The purpose was to concentrate on the drivers who had experienced accidents and investigate their characteristics and related factors in depth.
- Number of times: The average number of accidents experienced by the respondents is approximately 1.77, with a standard deviation of 1.16. The minimum and maximum values are 1 and 6, respectively.
- Location of the crash: This variable represents the location of the crash on a scale from 0 to 4, indicating different crash locations. The mean of 1.83 suggests that, on average, crashes occurred at location 1.
- Severity: The severity of the accidents is measured on a scale from 0 to 1, with 0 indicating non-severe accidents (only vehicle or property damage) and 1 representing severe accidents involving human injuries or deaths. The mean severity is approximately 0.26, indicating that, on average, accidents were more on the non-severe side.
- Causes: The variable represents the causes of accidents and ranges from 1 to 6. The mean of 3.17 suggests that, on average, accidents were caused by a variety of factors.

The methodology described above aimed to comprehensively analyze traffic accident data, integrating both statistical analysis and spatial visualization techniques. The combination of questionnaire-based data collection, decision tree analysis, and spatial analysis provided valuable insights into the patterns and determinants of traffic accidents in the study area.

3. Results and analysis

3.1. Road traffic accidents

The study's findings indicate that a significant proportion, specifically more than 45%, of the survey participants had been involved in at least one traffic accident. **Figure 2a** visually represents this information, depicting that nearly half of the respondents reported being involved in a car accident. Additionally, **Figure 2b** illustrates the distribution of participants who encountered a car accident within the city based on their respective positions inside the vehicle at the time of the accident.

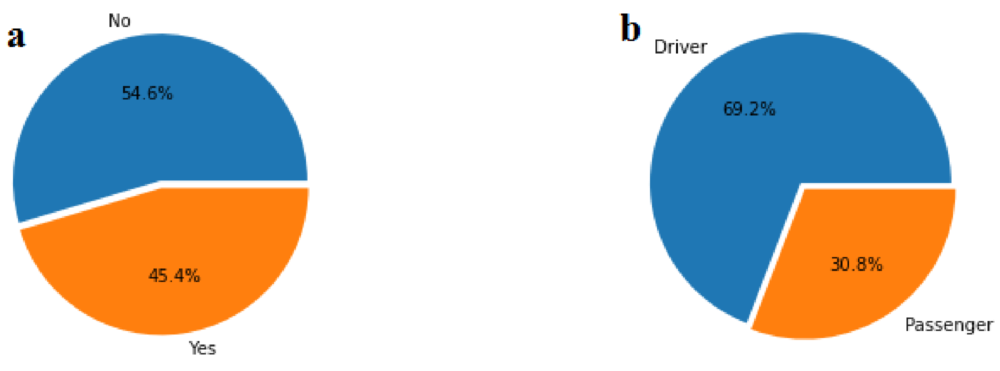

Figure 2. (a) individuals who had been in a road traffic accident; **(b)** position of the respondents who experienced car accidents.

3.2. Severity of the accidents

The study investigated the severity of crashes by assessing the level of human injuries, losses, and physical damage to the vehicles involved. The results reveal that more than 25% of the recorded accidents were categorized as severe crashes. This information is visually represented in **Figure 3**.

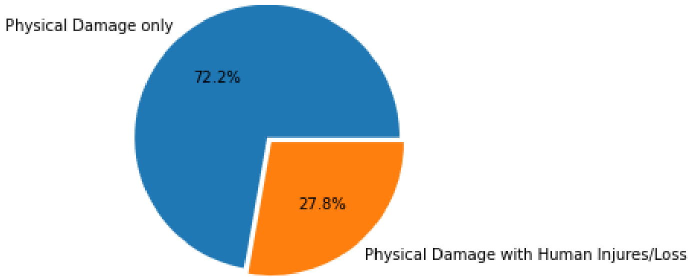

Figure 3. Percentage of severe crashes.

3.3. Having multiple accidents

Figure 4a presents the distribution of survey respondents based on their involvement in road accidents, highlighting that more than 40% of drivers have encountered multiple accidents. Additionally, **Figure 4b** provides the percentage of drivers who have experienced one or multiple accidents, while **Figure 4c** displays the percentage of multiple accidents categorized by gender.

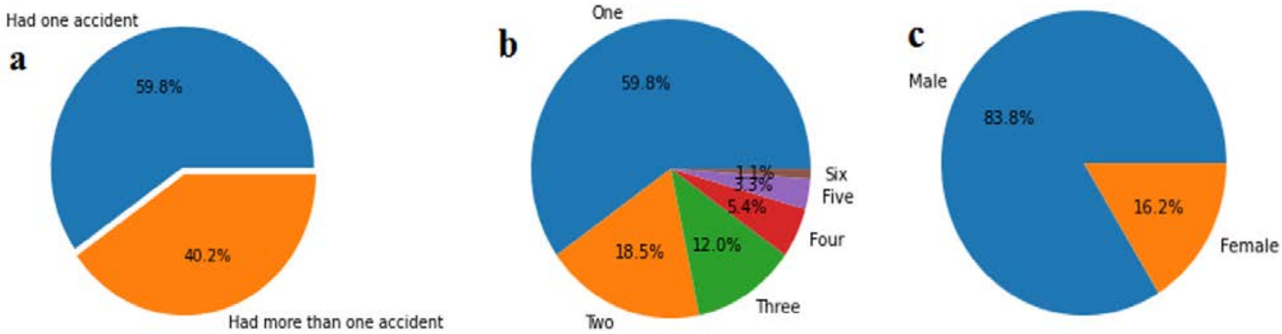

Figure 4. (a) drivers had been in one or more than one accident; **(b)** number of accidents reported by drivers; **(c)** male and female in multiple accidents.

3.4. Spatial analysis

In this section, Quantum GIS software was utilized for conducting spatial analysis. Specifically, the researchers employed the heat map package, an integrated library within the software, to generate heat maps of the city by entering the georeferenced longitude and latitude data of traffic accidents. The resulting heat map, presented in **Figure 5** using the first spatial map, depicted the concentration of all accidents that occurred within the city, highlighting areas of high occurrence, particularly in the CBD region. On the other hand, the second spatial map in **Figure 6** exclusively showed the locations of severe accidents, which were predominantly situated along highways and arterial roads.

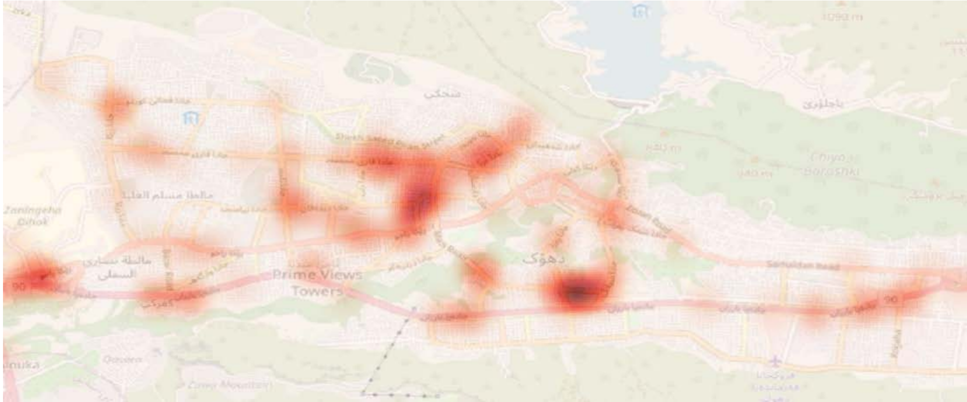

Figure 5. Heat map showing the spatial location of the hot spots of the traffic accidents in the city.

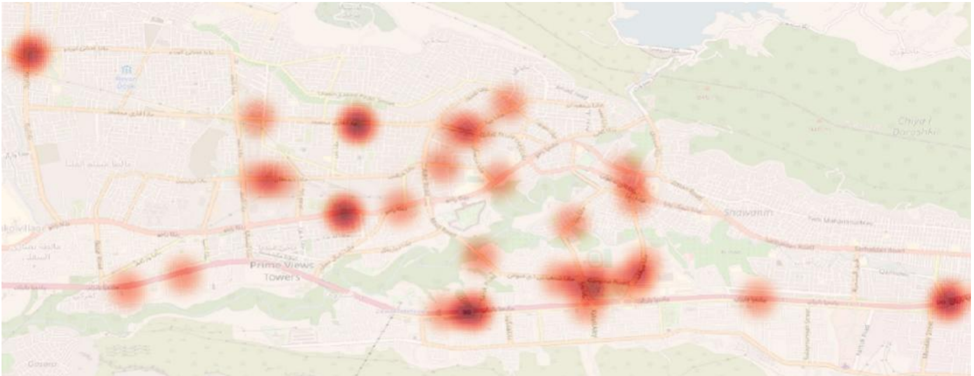

Figure 6. Heat map indicating the hot spots of the sever crashes in the city.

3.5. Decision tree analysis

The analysis of the decision tree has produced several significant insights. As can be seen in **Figure 7**, the variable "Causes" emerges as the primary influential factor in classifying objects, occupying the top position in the tree structure. The root contains the decision factor, entropy, samples, and an array value representing two numbers corresponding to the target variable, "multiple accidents," as outlined in **Table 2**. For instance, the array indicates that out of the sample, 220 drivers experienced one accident, while 148 drivers encountered more than one accident.

The left branches of the tree encompass accidents attributed to the first three causes: traffic conditions and road geometry, traffic law violations, and over speeding. These causes account for over 67% of the total accidents. In contrast, the right branches represent accidents resulting from causes indicated in **Table 2**, specifically causes numbered 4 to 6. Among these causes, the age factor is associated with crashes related to over speeding and traffic law violations. Notably, the youngest age group reports 33% of multiple accidents, while the older age groups report 100% of single accidents on local roads.

Conversely, the gender factor is situated in the right branch, which encompasses accidents caused by vehicle malfunctioning, drivers' behavior, and other factors such as climate conditions or road facilities. Within this branch, the proportion of multiple accidents is slightly above 50%. Furthermore, across all accident locations except the central business district (CBD), females account for a mere 15% of accidents, and among them, only 25% are classified as multiple accidents. In contrast, males account for the majority of accidents, comprising 85%, with approximately 70% of them being multiple accidents.

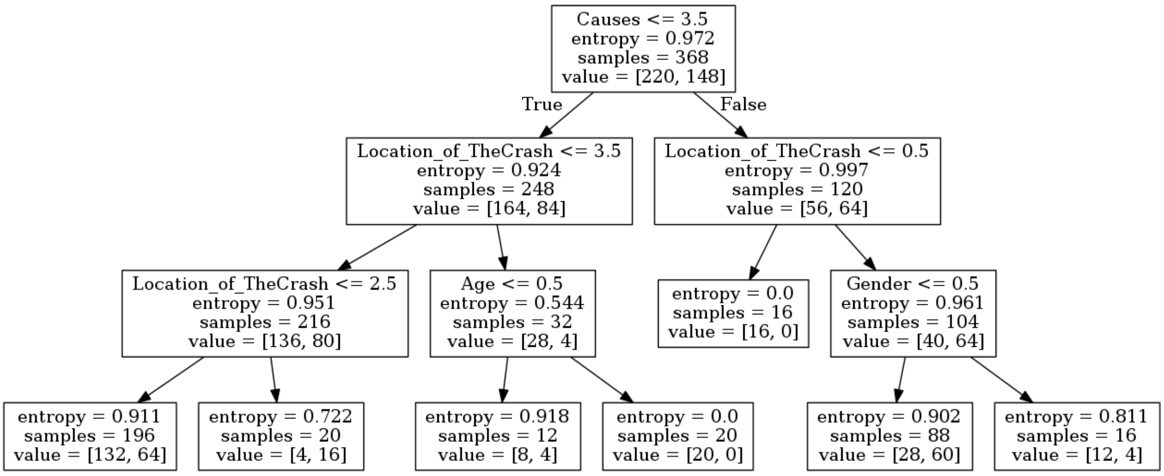

Figure 7. Python output displaying the DT method's graph.

This analysis yielded an accuracy rate of 71.73%, which serves as a strong validation of the decision tree (DT) method's efficacy in identifying the key factors contributing to multiple accidents' occurrence[20]. This high accuracy rate demonstrates the DT method's capability to effectively discern and predict the primary factors influencing the frequency of multiple accidents reported by drivers. The substantial accuracy rate signifies the reliability and precision of the model's predictions. Thus, the DT method proves to be a valuable tool for analyzing and comprehending the various factors that impact the occurrence of multiple accidents, thereby offering crucial insights for addressing road safety issues. The successful validation of the model affirms the credibility of the analysis conducted using the decision tree method.

While the study provides recognized insights into traffic accident patterns and socio-economic determinants, it is important to acknowledge the limitations:

- The study focuses on a specific city in the Kurdistan region of Iraq. The findings may not be applicable or generalizable to other cities or regions with different socio-economic characteristics or road infrastructure.
- The data collected through questionnaires rely on self-reported information from the participants. This introduces the possibility of response bias, inaccurate recall, or subjective interpretations, which could impact the reliability of the findings.

4. Conclusion and recommendations

This study examined the spatial distribution and socio-economic determinants contributing to traffic accidents in the city of Duhok, Kurdistan Region of Iraq, employing a mixed methodology that integrated the decision tree algorithm with Quantum GIS for spatial analysis. The findings revealed that the decision tree algorithm effectively identified key factors associated with multiple accidents, emphasizing the significance of traffic conditions, traffic law violations, and over speeding as major contributors to accident occurrence.

The practical implications of this research are noteworthy, as the identified influential factors can inform the development of targeted accident prevention strategies and road safety measures. Traffic authorities and urban planners can utilize these insights to implement measures aimed at improving road conditions, enforcing traffic regulations, and promoting safer driving behaviors to reduce accident rates in Duhok and potentially in other urban contexts.

The study's exploration of socio-economic determinants provides valuable insights into the underlying factors influencing accident occurrence. Addressing these determinants, such as driver education, behavioral interventions, and economic disparities, can significantly contribute to enhancing road safety and reducing accident rates.

Nevertheless, certain limitations should be acknowledged, including the focus on a specific city context, which may limit the generalizability of the findings to other regions. Moreover, the use of self-reported data through questionnaires introduces potential biases and accuracy concerns.

Despite these limitations, this research contributes to the field by integrating machine learning techniques and spatial analysis into traffic accident studies. The mixed methodology approach enriches the understanding of accident patterns and influential factors, distinguishing this study from previous works.

The study's results offer practical guidance for developing targeted interventions and policies to improve road safety. With traffic accidents remaining a pressing concern in modern cities, the insights and methodologies presented here can serve as a basis for further research and policy formulation, ultimately contributing to the creation of safer and more sustainable urban transportation systems.

Author contributions

Conceptualization, PA and TS; methodology, PA; software, PA; validation, PA; formal analysis, PA; investigation, PA; resources, PA; data curation, PA; writing—original draft preparation, PA; writing—review and editing, PA; visualization, PA; supervision, TS; project administration, PA; funding acquisition, PA. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

BME (Faculty of Transportation Engineering and Vehicle Engineering) and UoD (College of Spatial Planning) supported the research.

Conflict of interest

The authors declare no conflict of interest.

Nomenclature

QGIS: Quantum Geographic Information System

DT: Decision Tree

CBD: Central Business District

References

1. Kaparias I, Liu P, Tsakarestos A, et al. Predictive road safety impact assessment of traffic management policies and measures. *Case Studies on Transport Policy* 2020; 8(2): 508–516. doi: 10.1016/j.cstp.2019.11.004
2. Bucsuházy K, Matuchová E, Zůvala R, et al. Human factors contributing to the road traffic accident occurrence. *Transportation Research Procedia* 2020; 45: 555–561. doi: 10.1016/j.trpro.2020.03.057
3. Zefreh MM, Torok A. Theoretical comparison of the effects of different traffic conditions on urban road environmental external costs. *Sustainability* 2021; 13(6): 3541. doi: 10.3390/su13063541
4. Borsos A. Application of bivariate extreme value models to describe the joint behavior of temporal and speed related surrogate measures of safety. *Accident Analysis & Prevention* 2021; 159: 106274. doi: 10.1016/j.aap.2021.106274
5. Chand A, Jayesh S, Bhasi AB. Road traffic accidents: An overview of data sources, analysis techniques and contributing factors. *Materials Today: Proceedings* 2021; 47: 5135–5141. doi: 10.1016/j.matpr.2021.05.415
6. McKenzie G. Spatiotemporal comparative analysis of scooter-share and bike-share usage patterns in Washington, DC. *Journal of Transport Geography* 2019; 78: 19–28. doi: 10.1016/j.jtrangeo.2019.05.007

7. Ouni F, Belloumi M. Spatio-temporal pattern of vulnerable road user's collisions hot spots and related risk factors for injury severity in Tunisia. *Transportation Research Part F: Traffic Psychology and Behaviour* 2018; 56: 477–495. doi: 10.1016/j.trf.2018.05.003
8. Ouni F, Hzami M, Harizi R. Spatio-temporal hotspots analysis of pedestrian-vehicle collisions in Tunisian coastal regions. In: Proceedings of the 2020 IEEE 13th International Colloquium of Logistics and Supply Chain Management (LOGISTIQUA); 2–4 December 2020; Fez, Morocco. pp. 1–6.
9. Turki Z, Ghédira A, Ouni F, Kahloul A. Spatio-temporal analysis of road traffic accidents in Tunisia. In: Proceedings of 2022 14th International Colloquium of Logistics and Supply Chain Management (LOGISTIQUA); 25–27 May 2022; El Jadida, Morocco. pp. 1–7.
10. Belloumi M, Ouni F. Factors affecting the severity of motor vehicle traffic crashes in Tunisia. *SAE International Journal of Transportation Safety* 2019; 7(2): 99–128. doi: 10.4271/09-07-02-0006
11. Harizi R, Ouni F, M'rihi R. Detection and classification of road accident black zones using exploratory spatial data techniques. *International Journal of Trend in Research and Development* 2006; 3(1): 173–180.
12. Dereli MA, Erdogan S. A new model for determining the traffic accident black spots using GIS-aided spatial statistical methods. *Transportation Research Part A: Policy and Practice* 2017; 103: 106–117. doi: 10.1016/j.tra.2017.05.031
13. Sun TJ, Liu SJ, Xie FK, et al. Influence of road types on road traffic accidents in northern Guizhou Province, China. *Chinese Journal of Traumatology* 2021; 24(1): 34–38. doi: 10.1016/j.cjtee.2020.11.002
14. Kurakina E, Kravchenko P, Brylev I, Rajczyk J. Systemic approach to auditing road traffic accident black spots. *Transportation Research Procedia* 2020; 50: 330–336. doi: 10.1016/j.trpro.2020.10.039
15. Taamneh M. Investigating the role of socio-economic factors in comprehension of traffic signs using decision tree algorithm. *Journal of Safety Research* 2018; 66: 121–129. doi: 10.1016/j.jsr.2018.06.002
16. Clarke DD, Forsyth R, Wright R. Machine learning in road accident research: decision trees describing road accidents during cross-flow turns. *Ergonomics* 1998; 41(7): 1060–1079. doi: 10.1080/001401398186603
17. Bordarie J. Predicting intentions to comply with speed limits using a 'decision tree' applied to an extended version of the theory of planned behaviour. *Transportation Research Part F: Traffic Psychology and Behaviour* 2019; 63: 174–185. doi: 10.1016/j.trf.2019.04.005
18. da Cruz Figueira A, Pitombo CS, Meira E Silva de Oliveira PT, Larocca APC. Identification of rules induced through decision tree algorithm for detection of traffic accidents with victims: A study case from Brazil. *Case Studies on Transport Policy* 2017; 5(2): 200–207. doi: 10.1016/j.cstp.2017.02.004
19. Abellán J, López G, de Oña J. Analysis of traffic accident severity using Decision Rules via Decision Trees. *Expert Systems with Applications* 2013; 40(15): 6047–6054. doi: 10.1016/j.eswa.2013.05.027
20. Abdullah P, Sipos T. Drivers' behavior and traffic accident analysis using decision tree method. *Sustainability* 2022; 14(18): 11339. doi: 10.3390/su141811339
21. Abdullah P, Sipos T. Scenarios for mobility improvement based on individual socioeconomic factors. *The Journal of Duhok University* 2023; 26(1): 22–28. doi: 10.26682/sjuod.2023.26.1.3
22. Sobral T, Galvão T, Borges J. Visualization of urban mobility data from intelligent transportation systems. *Sensors* 2019; 19(2): 332. doi: 10.3390/s19020332
23. Fan Z, Liu C, Cai D, Yue S. Research on black spot identification of safety in urban traffic accidents based on machine learning method. *Safety Science* 2019; 118: 607–616. doi: 10.1016/j.ssci.2019.05.039
24. Abdullah PH, Perschon J, Ameen AM. The relationship between car dependency and use of public transport in Duhok city-barriers analysis and recommendations. *Journal of University of Duhok* 2020; 23(2): 59–68. doi: 10.26682/sjuod.2020.23.2.7