Engineered Science
DOI: https://dx.doi.org/10.30919/es1472

Spatial-Temporal Analysis and Severity Prediction of Traffic Accidents

Aslam Al-Omari,[1,*] Mohamed Alsaleh[1] and Nawras Shatnawi[2]

Abstract

Traffic accidents pose a significant problem, resulting in substantial human and economic losses annually. As such, safety engineers rely on studies of traffic accidents to implement remedial measures that prevent accidents or mitigate their impact on lives and the economy. This study explored spatial and temporal traffic accident patterns and developed a data-driven traffic accident severity prediction model. The traffic accident data used in the study were acquired for the City of Irbid, which is located in Jordan, covering the period from January 2016 to December 2021. The accident severity prediction model was developed using a Random Forest machine learning classification algorithm, and its optimal architecture was identified using a Bayesian optimization approach. Additionally, a model interpretability method, SHapley Additive exPlanations (SHAP), was employed to examine the impacts of the explanatory variables and their relative importance. The resulting model showed outstanding performance, with an overall predictive accuracy of 90%. Furthermore, the SHAP analysis results revealed that accident speed had the greatest influence on the model's predictions, followed in order by accident type, date, and primary road class. The results also indicated that higher accident speed values, collisions, and Run-Off-Road accident types, and downgrade curves are associated with higher accident severity classes.

Keywords: Accident severity, Random forest, Hotspot analysis, Traffic accidents, Machine learning.
Received: 07 December 2024; Revised: 21 February 2025; Accepted: 17 March 2025.
Article type: Research article.

1. Introduction

As per the World Health Organization (WHO) Global Status Report on Road Safety, an alarming 1.35 million fatalities were attributed to road traffic accidents in 2016.[1] Traffic accidents rank as the primary cause of death among children and young adults aged 5-29, according to the same report. This surpasses even HIV/AIDS and diarrheal diseases, standing in the eighth position as the leading cause of death across all age groups. Notably, the WHO report emphasized the significant variation in death rates between low-income and high-income countries, with the former experiencing three times higher fatality rates.

Given these worrying statistics, comprehensive studies on traffic accidents become crucial to conduct, particularly in low- to middle-income countries, to mitigate the global toll of fatalities. Jordan is a country situated in the Middle East (Asia). It falls within the category of middle-income countries, with a gross national income per capita of $3,920 in 2016.[1] The annual reports on traffic accidents from the Traffic Department of the Public Security Directorate (PSD) between 2016 and 2021[2-7] revealed an average economic loss of 314 million JDs per year, accumulating to a total of 1.884 billion JDs (1JD = 1.41USD) over the specified period. Moreover, they highlighted that approximately 3,700 individuals lost their lives in Jordan during the same period, averaging 616.5 deaths per year. Conducting accident analysis serves the purpose of identifying the cause or causes of an accident, thereby enabling preventive measures to avert similar incidents. The causes of traffic accidents can be classified into three principal categories: driver-related factors, vehicle-related factors, and road and environment-related factors.[8,9] These categories often influence traffic accidents in complex and overlapping ways. Exploring the severity of traffic accidents provides valuable insights into the interrelationships between these three categories and their impact on the severity of accidents.

[1] *Department of Civil Engineering, Jordan University of Science and Technology, Irbid, 22110, Jordan*
[2] *Department of Civil Engineering, Al-Balqa Applied University, Al-Salt, 19117, Jordan*
Email: aaalomari3@just.edu.jo (A. Al-Omari)

Furthermore, the development of prediction models for traffic accident severity is of vital importance to various stakeholders, including transportation safety planners, hospitals, emergency care providers, and insurance companies.[10] Additionally, studying traffic accidents and identifying hotspot locations offers invaluable assistance to decision-makers in both the pre-remediation phase, where necessary countermeasures are devised to establish a robust traffic transportation system,[11] and the post-remediation phase, which involves evaluating the effectiveness of implemented measures.[12] These insights empower decision-makers to make informed choices and continually improve road safety measures. With the ongoing expansion of the population, there has been a significant increase in the number of automobiles in recent times, necessitating a thorough examination of factors influencing accident severity. The goal is to identify key elements that exacerbate the losses associated with traffic accidents. A comprehensive understanding of these factors is crucial for the implementation of effective measures to reduce accident occurrences and mitigate their severity, ultimately leading to minimized economic losses and saved lives. The following subsections offer a background on both the analysis of accidents in a spatiotemporal context and the development of models for predicting accident severity.

To understand the spatial distribution of accidents, geographic information system (GIS) tools have proven to be invaluable. Using GIS, Ivan and Haidu highlighted the concentration of accidents in the central area of Cluj-Napoca.[13] The concentration was attributed to increased pedestrian flow and a high density of intersections in this urban region. Furthermore, these authors conducted a Hotspot Analysis to identify spatial clustering patterns, distinguishing between “hot spots” with high values and “cold spots” with lower values. The authors employed Kernel density analysis to determine the gravity of accidents within a 500 square meter radius, taking into account the number of “victims”. This method, known for its simplicity and ease of use, offers valuable insights into spatial data analysis.[14,15] Moreover, Al-Omari *et al*.[11] used GIS and fuzzy logic analysis to identify hotspots in traffic accidents in Irbid from 2013 to 2015. The weighted overlay method (WOM) and fuzzy overlay method (FOM) were employed to predict accident hotspots. They revealed certain streets and intersections that exhibited elevated accident frequencies. Factors such as volume rates, speeds, heavy vehicle percentages, and road geometry contributed to these increased risks, highlighting the value of GIS analysis and fuzzy logic in understanding accident hotspots.

This analysis sheds light on the geographic patterns of accidents and their underlying causes, which are central to understanding and addressing road safety. Since the outcomes of the spatio-temporal analysis can be effectively employed by various organizations to enhance traffic planning and management, ultimately leading to improved traffic conditions and a reduction in accidents.[16]

When it comes to the accident severity prediction models, machine learning methods have emerged as effective alternatives to overcome the limitations of conventional statistical methodologies due to their modeling flexibility, generalization ability, and impressive predictive performance.[17] Various machine learning algorithms, including random forest, K-nearest neighbor, Naive Bayes, and others, have been widely employed to address the complexities of traffic safety analysis and prediction tasks. Malik *et al*.[18] predicted accident severity using crash datasets, as they compared six machine learning algorithms: Random forest (RF), logistic regression (LR), decision tree (DT), Naïve bayes (NB), AdaBoost, and bagging. Notably, RF, bagging, and DT outperformed other machine learning algorithms, achieving superior results for all performance metrics. Accident severity was divided into three classes: slight, serious, or fatal. Similarly, evaluating the performance of different ensemble machine learning methods on road accidents, Ahmed *et al*.[19] reported that the RF classifier outperformed other methods, as evidenced by the area under the receiver operator characteristic (AUROC) and other performance metrics. A hybrid approach to building a three-class accident severity prediction model was employed by Yassin and Pooja.[20] Utilizing K-means clustering and RF classification to predict road accident severity and identify contributing factors, their approach achieved an exceptional accuracy of 99.86%. They identified that light condition, driver's age and experience, and vehicle service year were strongly related to the prediction of accident severity. In addition, Geyik and Kara adopted four machine learning algorithms to model traffic accident severity,[21] including multilayer perceptron (MLP), DT, RF, and Naive Bayes classifiers. The results revealed overall accuracies of 85.19% and 86.67% for the RF and MLP models, respectively, outperforming other models. In Jordan, Al-Mistarehi *et al*.[22] analyzed and developed prediction models for accident severity in Al-Zarqa City using data from 2014 to 2018. The authors employed association rules, an unsupervised learning approach, to identify factors influencing accident severity. Moreover, DT, RF, and boosted tree algorithms were used to model accident severity levels. Amongst all, the RF model provided the highest predictive accuracy.

Furthermore, RF, as an improved version of the decision tree algorithm, was found to address the issue of overfitting and offer benefits such as high prediction accuracy, efficient handling of large datasets, and the ability to manage unbalanced and missing data.[23] As for the hyperparameter tuning for machine learning models, Turner *et al*.[24] demonstrated the superior efficiency of Bayesian optimization over random search. Jadaan *et al*.[25] provided countermeasures that exhibit the highest favorability and effectiveness for implementation, in Jordan, encompass the following key strategies: the provision of safe sidewalks for pedestrians, enhancement of road infrastructure, increased support for public transportation, and the spread of traffic awareness within educational institutions (schools).

Overall, it is important to note that drivers' behavior (*e.g.*, aggressive behavior and errors), cultural background, and experience in dealing with different weather conditions, pavement surface conditions, and road geometry/types differ from one country to another, which highlights the importance of developing country-specific models.[26-28] To the best of the authors' knowledge, there are no previous studies that have analyzed traffic accident hotspots or developed interpretable data-driven models to predict traffic accident severity level in Jordan. Specifically, in Irbid, which is the second-largest contributor to traffic accidents in Jordan. With this in mind, this study utilizes a comprehensive traffic accident data for the City of Irbid, Jordan, from January 2016 until December 2021 to a) explore spatial and temporal traffic accident patterns, b) identify hotspot locations, and c) establish an interpretable data-driven traffic accident severity prediction model. The data utilized in this study were acquired for Irbid from the Central Traffic Police Department (CTPD). The dataset encompasses a comprehensive set of information pertaining to traffic accidents, including details about accidents themselves, drivers and vehicles involved, road surface conditions, road geometric characteristics, and weather and visibility conditions.

2. Research framework

This section illustrates the study area as well as the research framework employed in this study, outlining the key steps followed to analyze spatial and temporal traffic accidents patterns, identify hotspot locations, and develop a prediction model for accident severity levels. This section provides a comprehensive overview of each of these steps.

2.1 Study area

The current study focuses on the examination of traffic accidents within Irbid City, situated in the northern region of Jordan. As indicated in Fig. 1, Irbid City stands as the second most populous urban center in Jordan, following the capital City of Amman. According to the data presented by the Department of Statistics (DOS) for 2021,[29] Irbid encompasses a total land area of 1572 square kilometers, accommodating a population of 2,050,300 residents in 2021. Consequently, Irbid boasts the highest population density in the country at a rate of 1304.4 individuals per square kilometer. The city's road network spans a combined length of 876.5 kilometers, comprising highways, secondary roads, and rural roads measuring 162, 380, and 334.5 kilometers, respectively. Moreover, Irbid has a total of 117,013 licensed vehicles, with 103,675 classified as private vehicles and 13,338 as public vehicles.[30,31]

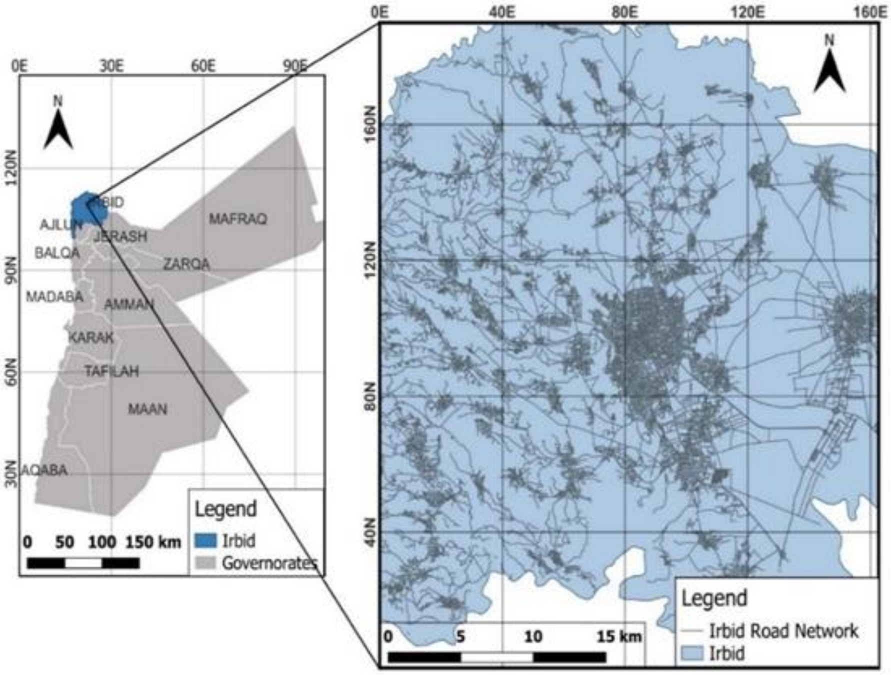

Fig. 1: Jordan map with major governorates and Irbid map with the road network.[31]

Table 1: Brief summary of accidents in Irbid in the period (2016-2021).

Measures	2016	2017	2018	2019	2020	2021	Sum	Source
Fatalities	93	95	78	75	60	70	471	[2-7]
Low severity Injuries	2676	2526	2021	1755	1450	2661	13089	
Medium severity Injuries	1	1	525	654	529	756	2466	
High severity injuries	252	196	98	81	76	92	795	
Property damage only (PDO)	12998	13936	14661	13507	11318	15342	81762	
Citizens in Irbid (×1000)	1819.6	1867.0	1911.6	1957.0	2003.8	2050.3	-	[29,30]
Number of registered vehicles	117013	116825	116399	125973	138300	141754	-	[2-7]
% Increase	-	-0.16	-0.364	+8.22	+9.78	+2.49	-	

Notably, following the capital city, Irbid exhibited the second-highest contribution to the occurrence of injury accidents in Jordan, accounting for an average rate of 16.6% over the study period of 2016-2021, as described in the annual accident reports by the PSD, as shown in Table 1. The study dataset disclosed that during the six-year duration, 471 lives were tragically lost due to road accidents in Irbid, averaging 78.5 deaths per year. Furthermore, a total of 16,300 injury accidents of varying severity, ranging from low to high, were recorded.[2-7]

2.2 Data collection

The study data were obtained from the CTPD located in Amman, which operates under the PSD. The study area encompassed the entire road network of Irbid city, including arterials, highways, secondary and rural roads, as well as major intersections. A total of 177,378 observations were collected, representing the number of vehicles involved in accidents over a six-year period from 2016 to 2021. The dataset contained vital information about each accident, covering the following information:

• Driver-related attributes, including age, gender, and driver errors (*e.g.*, running a red light, tailgating, not stopping at a stop sign, *etc*.).

• Road-based factors, including road type (*i.e.*, one-way, divided two-way, undivided two-way).

• Road surface type (*i.e.*, asphalt, aggregate, concrete, unpaved, and others) and conditions (*i.e.*, dry, wet, icy, oily, and others).

• Road geometry (*i.e.*, straight and level, straight with upgrade, straight with downgrade, curved and level, curved with upgrade, curved with downgrade).

• Weather (*i.e.*, clear, rainy, snowy, dusty, *etc*.) and visibility conditions (daylight, dark, night with insufficient light, night with sufficient light).

• Vehicle-related details comprised the vehicle's category and plate origin; and

• Accident-related characteristics including date and time, location (longitude and latitude), type (*i.e.*, collision, run-off-road, pedestrian), speed, and severity (*i.e.*, fatalities, injuries, and property damage only accidents (PDOs)).

This comprehensive dataset served as the foundation for the study, providing the necessary information to conduct the analysis.

2.3 Data preprocessing

Data preprocessing, also known as data cleaning, encompasses multiple stages. Preprocessing had to be executed carefully to ensure unbiased modifications to the dataset. The goal was to prepare the data for subsequent data processing operations. Consequently, this necessitated a comprehensive exploration of data science principles to carry out these modifications in a neutral and logical manner. The main steps followed in this study for data preprocessing are summarized as follows:

a) Data Auditing and Validation: This stage involves thoroughly examining the dataset. Identifying necessary modifications and determining the required data preprocessing steps to ensure they align with the expected data type and fall within the anticipated range of values.

b) Data Reduction: Involved removing duplicates, reducing the complexity, and removing unnecessary features.

c) Data Integration: Involved consolidating the 144 files into a single file that encompasses the entire dataset. The integration step is crucial for processing and modeling purposes, ensuring a consistent and clean dataset.

d) Data Formatting: Data formatting was applied to the time and date fields to achieve a desired and readable format suitable for the model's requirements.

e) Data Filtering: Removing irrational, unexpected values, and outliers resulting from data collection mistakes. For instance, negative values and illogical age ranges.

f) Data Imputation: Filling in missing data to ensure the usability of the dataset. Fields with blank entries or missing values were replaced with "Unknown Value" to avoid compromising the results or losing portions of the dataset.

g) Data Transformation: Data transformation was necessary for spatial analysis of accidents and modeling purposes. It entailed converting categorical data into numerical values.

2.4 Methods

The main methods and algorithms adopted in this study included kernel density estimation (KDE), RF, Bayesian optimization, and Shapley additive explanations (SHAP). Each method was selected for its high performance in fulfilling specific tasks aligned with the objectives of this study. KDE is a spatial analysis tool that builds a continuous surface, which facilitates the identification of areas with high accident concentrations (high-risk areas), visualized as a smooth density map.[32] For severity prediction, the RF algorithm was utilized. This method combines multiple DTs, making it robust against overfitting and effective at handling non-linear relationships between accident severity parameters, with high predictive accuracy in both regression and classification tasks.[33] To further enhance model accuracy, Bayesian optimization was employed for automatic hyperparameter tuning of the RF model. This approach is considered superior to traditional tuning methods, as it reduces computational cost while achieving optimal model performance. Finally, SHAP was used to interpret the predictions of the accident severity model, revealing how different accident variables (such as speed or weather conditions) influenced severity levels.[34] Together, these methods formed a robust analytical framework that effectively supported the objectives of this study.

2.4.1 Kernel density estimation

Kernel density estimation is a method that creates a smooth and continuous surface by placing circular areas, known as kernels, over observed points. These kernels, with a defined bandwidth, capture the indicator value at each point and spread it according to a suitable function. The sum of these values across all locations, even where no incidents were recorded, generates a density estimate surface. Density can be measured using two methods: the simple method and the kernel method. The simple method divides the study area into cells and calculates the density based on the ratio of features within circular neighborhoods drawn around each cell. The resulting density map is influenced by the radius of the circular neighborhoods, where a larger radius leads to a smoother surface by incorporating more feature points.[32] In contrast, the kernel method divides the study area into cells and assigns circular neighborhoods around each individual feature point, such as accidents. A mathematical equation is then applied that gradually decreases the indicator value from 1 at the feature point to 0 at the neighborhood boundary.[35]

2.4.2 Random forest

RF is an ensemble machine learning algorithm that is primarily based on decision trees. It employs the bagging technique to create multiple subsets of data by randomly selecting observations from the original dataset with replacement.[19] These subsets are then used to construct smaller trees. In RF, all these trees are run in parallel, and the final prediction is obtained by averaging the predictions from all the decision trees. Key parameters such as the number of trees, the maximum number of features considered for splitting a node, and the minimum number of leaves required to split an internal node play a crucial role in enhancing the prediction accuracy of RF.[19] RF algorithm is a widely utilized non-linear statistical ensemble method that finds extensive application in regression and classification tasks. Distinguished from standard decision trees, RF adopts a distinctive approach to node splitting. At each node, rather than considering all variables, RF selects the best split from a randomly chosen subset of predictors. Furthermore, RF resiliency is powerful against overfitting, establishing itself as a dependable choice for predictive modeling.[33]

2.4.3 Bayesian optimization

The primary objective of hyperparameter optimization is to automate the process of tuning hyperparameters, enabling users to effectively apply machine learning models to real-world problems.[36] It is important to conduct the process of tuning the hyperparameters of a particular machine learning model to find its optimal structure, leading to the optimal performance of the model. Tuning hyperparameters in the traditional approach involves setting and testing several combinations of hyperparameters manually. However, this method becomes inefficient as it takes much time when many hyperparameters are being evaluated. Automatic hyperparameter tuning approaches include grid search, random search, and Bayesian optimization. The Bayesian optimization approach involves manual creation of a hyperparameter space from which the algorithm can choose a set of hyperparameters for model training.[37] Then, a cross-validation score is returned by the trained model. The surrogate function uses the score returned to suggest new hyperparameters to the objective function, which is defined to improve the performance. The predefined number of iterations is achieved through the repetition of the iterative process. This automatic approach is more effective and efficient compared with other automatic tuning approaches, as it often takes fewer evaluations to find the optimal hyperparameter set.[36]

2.4.4 Shapley additive explanations

SHAP is an approach rooted in game theory that offers a means to explain the predictions made by machine learning models.[34] This is accomplished using a concept known as Shapley values, as it quantifies the influence and relative significance of each feature in relation to the model's output. Shapley values are derived by comparing the model's predictions with and without the presence of a specific feature, leading to assessing the impact of that feature on the predictions. The sign of these values indicates whether the corresponding feature has a positive or negative effect on the model's prediction. As per the magnitude, it reflects the extent of its contribution. As a result, feature importance is determined by aggregating the absolute Shapley values over the samples in the testing dataset.[38-40]

2.5 Descriptive analysis

Temporal and spatial analyses were conducted to gain insights into the distribution of accident features both in terms of time and location. This analysis helped establish a solid foundation for building the RF model, ensuring its accuracy and reliability.

2.5.1 Temporal distribution

A thorough temporal analysis was conducted to examine the occurrence of accidents across different time intervals. Using MS Excel, histograms were created to visualize the hourly, daily, monthly, and yearly patterns of traffic accidents. This approach provided insights into how accidents were distributed over various time periods, revealing temporal patterns and identifying peak accident occurrence times. Furthermore, histograms were generated to plot the total number of accidents that took place during the six-year study period, alongside various accident features. Line graphs were incorporated into these plots to represent the percentage of accidents associated with each element within the respective category being studied. Analyzing the distribution of different accident features in relation to the total number of accidents contributes to a deeper understanding of the dynamics surrounding traffic accidents within the city.

2.5.2 Spatial distribution

The spatial distribution of accidents in Irbid over the study period was examined using GIS. Initially, the total number of accidents was visually represented to identify concentrated areas of accidents within the city. The distribution of accidents was analyzed based on their severity, specifically focusing on PDO and fatality accidents. This analysis aimed to establish a deeper understanding of the relationship between the road network and types of accidents. To present the accident locations in Irbid, the coordinate reference system (CRS) ESPG: 32636 WGS 84 UTM Zone 36N was utilized. A kernel density-based heat map was developed to identify hotspot locations. Hotspots were identified based on the overall intensity of accidents, regardless of their severity level. A radius of 100 meters was assigned to this approach, providing hotspot locations at the link level. This enabled a graphical representation of accident patterns and hotspot locations in Irbid, offering valuable insights into the spatial distribution of accidents within the city.

2.6 Model specifications

Upon completion of data collection and preprocessing, the dataset acquired was utilized to model the severity of traffic accidents. When constructing a supervised machine learning model, a critical aspect involves identifying the explanatory variables (independent variables) and the dependent variables (target variables). In this regard, the severity of the accidents served as the basis for determining the dependent variable, as indicated explicitly in Table 2. It is noteworthy that the dependent variable encompasses five distinctive categories: PDO, simple injury, intermediate injury, severe injury, and fatality. The explanatory variables of the model were determined by utilizing accident-related data, road-related data, and weather-related data. From the accident-related data, six important features were defined: the type of accident, the speed involved, the year it happened, the month, the particular day of the week, and whether the accident occurred on a weekend. In relation to road-related data, a total of seven explanatory variables were derived. These variables encompassed road type, surface type, surface condition, road geometry, lighting condition, road link class, and road class. As for the weather condition-related data, only one variable, namely weather condition, was incorporated. Table 2 provides the definitions and specifications for both dependent and independent variables.

The resulting dataset contained 91,950 accidents, with 81,772 PDOs, 7,056 low injury, 2,047 intermediate injury, 658 high injury, and 417 fatality accidents. It can be observed that the dataset is highly unbalanced, which can affect the performance of the model. Therefore, the over-under sampling technique, SMOTETomek algorithm, has been used to perform over-sampling for under-represented classes and

Table 2: Definition of target, accident-related, road-related, and weather-related variables.

Variable	Definition	Variable	Definition
Target (Dependent) Variables			
Accident Severity	= 0 for PDO = 1 for Low Injury = 2 for Intermediate Injury	Accident Severity	= 3 for High Injury = 4 for Fatality
Accident-related Variables			
Accident type	= 0 for Pedestrian Accident = 1 for Collision = 2 for ROR	Vehicle speed	Values used as is, ranging from 10 to 120 (km/h)
Year	= 1 for 2016 up to = 6 for 2021	Month	= 1 for January up to = 12 for December
Day of week	= 1 for Sunday up to = 7 for Saturday	Weekend	= 1 for Fri. and Sat. = 0 for other weekdays
Road-related Variables			
Surface type	= 0 if the surface is asphalt = 1 for concrete = 2 for aggregate = 3 for unpaved surfaces	Lighting condition	= 0 for daylight, sunrise, sunset = 1 for night with enough light = 2 for night with insufficient light = 3 for dark lighting conditions
Surface type	= 0 if the surface is asphalt = 1 for concrete = 2 for aggregate = 3 for unpaved surfaces	Lighting condition	= 0 for daylight, sunrise, sunset = 1 for night with enough light = 2 for night with insufficient light = 3 for dark lighting conditions
Road class			
Primary	= 1 if primary road = 0 otherwise	Service	= 1 if service road = 0 otherwise
Secondary	= 1 if secondary road = 0 otherwise	Construction	= 1 if construction road = 0 otherwise
Tertiary	= 1 if tertiary road = 0 otherwise	Residential	= 1 if residential road = 0 otherwise
Trunk	= 1 if trunk road = 0 otherwise	Unclassified	= 1 if unclassified road = 0 otherwise
Road geometry			
Straight & level	= 1 for straight & level, = 0 otherwise	Curved & level	= 1 for curved & level, = 0 otherwise
Straight & upgrade	= 1 for straight & upgrade, = 0 otherwise	Curved & upgrade	= 1 for curved & upgrade, = 0 otherwise
Straight & downgrade	= 1 for straight & downgrade, = 0 otherwise	Curved & downgrade	= 1 for curved & downgrade = 0 otherwise
Road link class			
Primary	= 1 if primary link = 0 otherwise	Tertiary	= 1 if tertiary link = 0 otherwise
Secondary	= 1 if secondary link = 0 otherwise	Trunk	= 1 if trunk link = 0 otherwise
Weather-related variables			
Weather condition	= 0 for Clear weather = 1 for Rainy		= 2 for Icy = 3 for Foggy and Dusty

under-sampling for over-represented classes simultaneously. Following the application of the over-under-sampling technique, the modelling dataset contained 407,482 instances distributed almost evenly among the severity classes, with approximately 81500 instances per class.

RF algorithm has been employed to model the traffic accident severity classes. The modelling dataset has been divided into training and testing with a 4:1 ratio, which is a commonly used split ratio for machine learning models.[41] A Bayesian optimization approach was employed to determine the optimal hyperparameter configuration for the algorithm. A search space was defined, consisting of various hyperparameters including Criterion, Maximum depth, Maximum features, Number of estimators, Minimum samples split (Int), and Minimum samples leaf (Int). In order to conduct the optimization, the tree parzen estimator (TPE) algorithm was utilized as the surrogate function, and the maximum number of iterations was set to 30. Each iteration involved training models and evaluating their performance using a 5-fold cross-validation strategy.

The objective of the optimization was to minimize the negative accuracy score, aiming to identify the hyperparameter set that would yield the highest cross-validation score. The resulting optimal hyperparameter configurations are presented in Table 3. The hyperparameter “Criterion” allocates the function used to evaluate the quality of a split, and it was found that the optimal choice is Gini. The "Maximum depth" refers to the maximum depth allowed for the tree, and it was found that a depth of 30 yields the best results. Regarding the "Maximum features," which determines the number of features to consider for each split, it was observed that using the square root of the total features is the optimal approach. In terms of the number of trees in the model (number of estimators), a value of 10 was identified as the most effective. Lastly, for both the "Minimum sample split" and "Minimum samples leaf," which indicate the minimum number of samples required for splitting an internal node and being at a leaf node, respectively, the optimal values were determined to be 3 and 1. Moreover, the optimal architecture of the algorithm was applied to the testing set to find the prediction accuracy. The performance of the final model was assessed using accuracy, precision, recall, and F1 score metrics described by Ahmed *et al*.[19] Finally, SHAP was employed to explain the results and capture the importance of the explanatory variables.

Table 3: Optimal hyperparameter set for the accident severity model.

Hyperparameter	Optimal set
Criterion	Gini
Maximum depth	30
Maximum features	Sqrt
Number of estimators	10
Minimum samples split (Int)	3
Minimum samples leaf (Int)	1

3. Results and discussion

Among 91950 accidents covering a period of six years, a substantial 95.5% of these accidents occurred primarily under clear weather conditions. Additionally, 94.4% of accidents transpired on straight and level roads. Further investigation revealed that collision accidents dominated the accident categories, accounting for 94.4% of total incidents. These collisions had the highest fatality rate, causing 218 fatalities during the study period, followed by pedestrian accidents with 4.3% and 1.3% for run-off-road (ROR) accidents. In terms of roadway types, the majority of accidents occurred on divided two-way roads, accounting for 54% of the total, while undivided two-way roads were the site of 37% of the accidents. It is worth noting that undivided two-way roads had a higher percentage of fatal accidents compared to other road types. Regarding accident speeds, the majority occur within the range of 40 to 60 km/hr. This trend can be attributed to the fact that most accidents transpire within city limits, where speed limits commonly fall within this range. Impressively, 91% of accidents occurred at 60 km/hr or lower, and among these, accidents at 60 km/hr exhibited the highest fatality rate, primarily reinforcing the importance of road safety measures within city limits. This section presents the results related to the spatial and temporal patterns, hotspot locations, and predictive model for accident severity levels.

3.1 Temporal and spatial patterns

3.1.1 Temporal patterns

Fig. 2 shows the yearly, monthly, daily, and hourly distributions of accidents. According to Table 1 from Jordan's Department of Statistics yearbooks for 2016 and 2021,[29,30] the number of registered vehicles in Irbid experienced the highest increase rate in the year 2020, reaching 138,300 registered vehicles in the city, which corresponds to a growth rate of 9.78%. However, Fig. 2a depicts the annual distribution of accidents, revealing that 2020 had the lowest number of accidents, with a total of 12,671 accidents, accounting for only 13.78% of the total number of accidents in the 6-year study period. This can be attributed to the COVID-19 pandemic and the subsequent strict measures implemented by Jordan, including lockdowns, which significantly restricted individuals' movements. This played a pivotal role in reducing the number of accidents in 2020. In contrast, despite a normal percentage increase of 2.49% in the number of registered vehicles in 2021, the year marked a substantial shift in accident statistics. Remarkably, 2021 recorded the highest percentage of accidents at 18.84%, with a total of 17,323 accidents reported in the city. This can be attributed to the aftermath of the strict pandemic measures, which seemed to

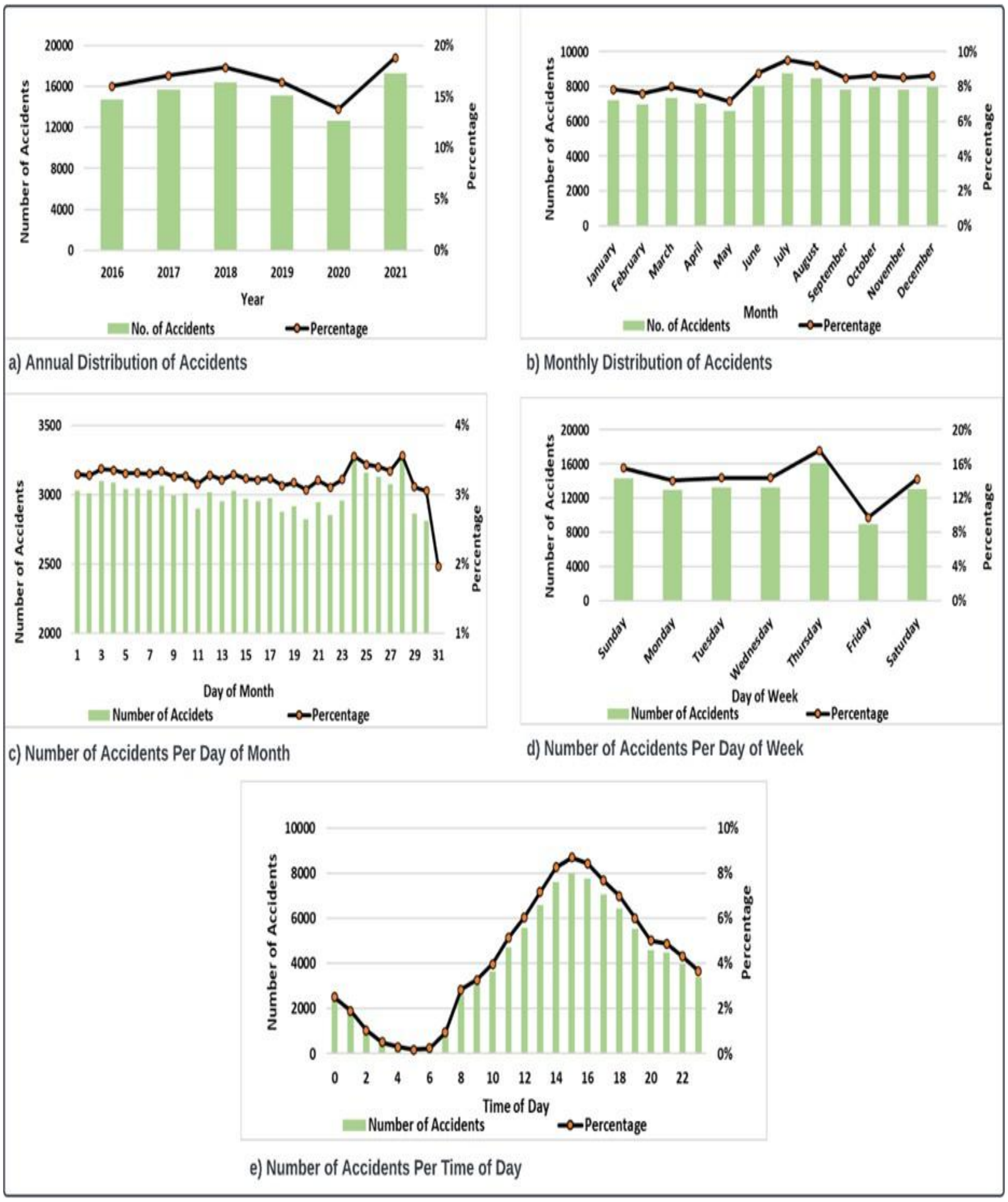

Fig. 2: Temporal analysis of traffic accidents.

have impacted the psychological state of residents, resulting in increased aggressive driving behaviors and subsequently leading to a rise in accidents.

According to Fig. 2b, the summer season, specifically June, July, and August, exhibits the highest rate of accidents. This can be attributed to the increased number of daily trips during the summer season. The higher traffic load during this period directly contributes to an escalation in accidents. Fig. 2c reveals that the last few days of each month observe a higher rate of accidents. Explained by the fact that people receive their paychecks at the end of the month, leading to a combination of increased economic activity and higher travel volumes during this time, resulting in elevated accident rates.

In Fig. 2d, it is evident that Thursday records the highest percentage of accidents at 18%. As the last workday of the week, Thursday sees a rise in the number of trips, including people leaving the city to return to their hometowns and vice versa. Sunday follows closely behind with a percentage of 16%, as it marks the beginning of the workweek and entails people traveling to and from other cities, such as university students returning to Irbid. Conversely, Friday registers the lowest accident percentage at 10% since it is an official holiday for both the public and private sectors, resulting in fewer commuting trips and subsequently a decreased number of accidents on this day. Lastly, Fig. 2e shows that the hours between 2 PM and 4 PM represent the peak period for accidents during the day. These hours coincide with the conclusion of official working hours for employees and students, leading to traffic congestion and increased traffic volume. Meaning a higher likelihood of accidents occurring during this time frame. Fig. 2 highlights various temporal patterns in accident occurrence, including annual, monthly, day of the week, and hourly trends. Understanding these patterns can help develop targeted strategies and interventions to enhance road safety during these critical periods.

3.1.2 Spatial distribution

Based on conducting spatial distribution using GIS on the study dataset, it was found that a significant portion of accidents fall within the PDO accident severity category, with a notable concentration in the central region of the city. The majority of accidents are categorized as low injury, followed by medium injury incidents, while high injury accidents are the least frequent. Moreover, when examining the distribution of 417 fatal accidents occurring over the six-year study period, the central region of the city emerged as the primary location for these tragic events. Illustrated in Fig. 3 is an overview of the road infrastructure in the city of Irbid. The figure highlights the presence of 31 key intersections located within the central region of the city, clearly indicated by cross marks.

Illustrated in Fig. 4 is the spatial distribution of all accidents, encompassing PDO accidents, injury occurrences, and fatality cases. Notably, regardless of their severity, accidents tend to be concentrated primarily in the central zone of the city. This concentration can be attributed to the central area's function as a pivotal hub for numerous commercial and business endeavors, attracting a substantial influx of individuals engaged in work, shopping, and related activities. Additionally, it serves as a nexus for major intersections, interconnecting different sections of the city. These collective factors contribute to escalated traffic congestion, ultimately resulting in the clustering of traffic accidents within the region. Accidents, irrespective of their classification, exhibit a notable concentration within the central region of the city.

3.2 Hotspot identification

To gain a deeper understanding of the spatial pattern of accidents, kernel density analysis was employed without any weighting. Based on Fig. 5, an unweighted kernel density heatmap reveals that the major arterials in the city center exhibit the highest kernel density values. Moreover, certain road segments can be identified as accident-prone areas due to their association with the highest kernel density values. These segments include Alhashmi section 1, all sections of King Abdulla II, sections 2 and 3 of Shafiq Irsheidat, sections 1 and 2 of Prince Hasan, Rateb Bataineh, all sections of King Husain, and section 1 of Umar Almokhtar.

3.3 Modelling

This subsection provides the results of the traffic accident severity level model that was developed using RF algorithm. The performance of the model on the testing dataset is presented in Table 4. Notably, the model consistently exhibited a high level of predictive accuracy, as indicated by precision, recall, and F1-score, across the five severity classes. Out of the total 15,692 instances of PDO, the model accurately predicted 13,726 instances, resulting in precision of 0.9, recall of 0.87, and F1-score of 0.89. Furthermore, the model correctly

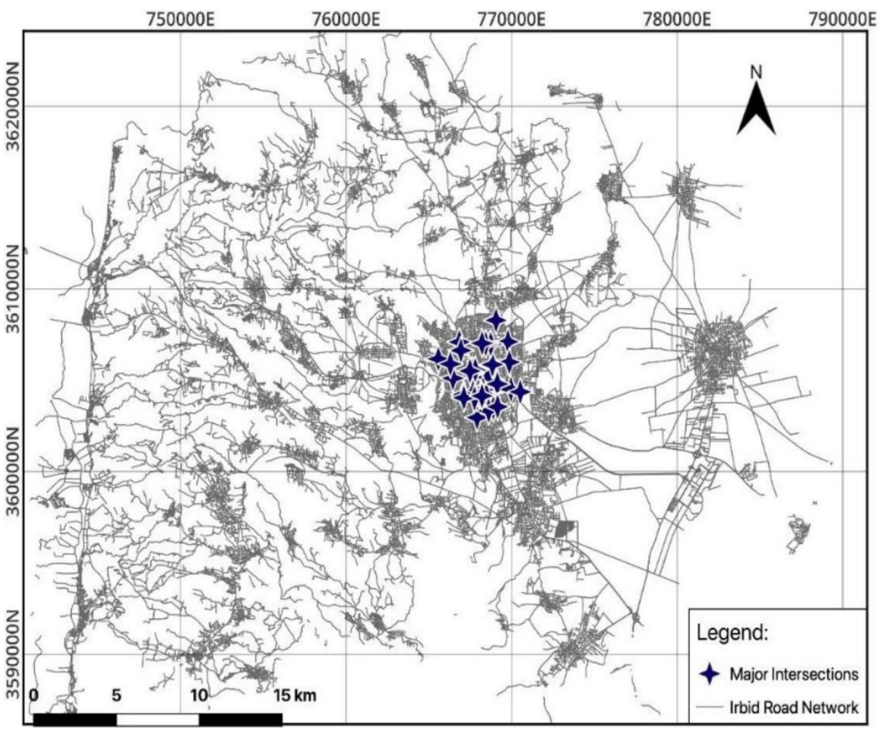

Fig. 3: Irbid road network and major intersections.

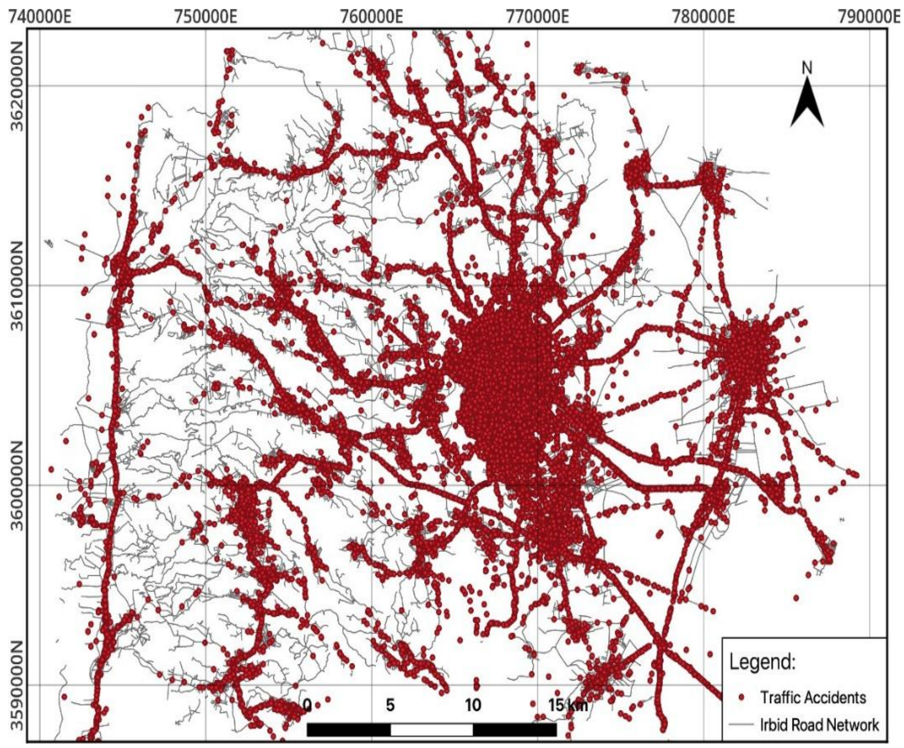

Fig. 4: Accident distribution.

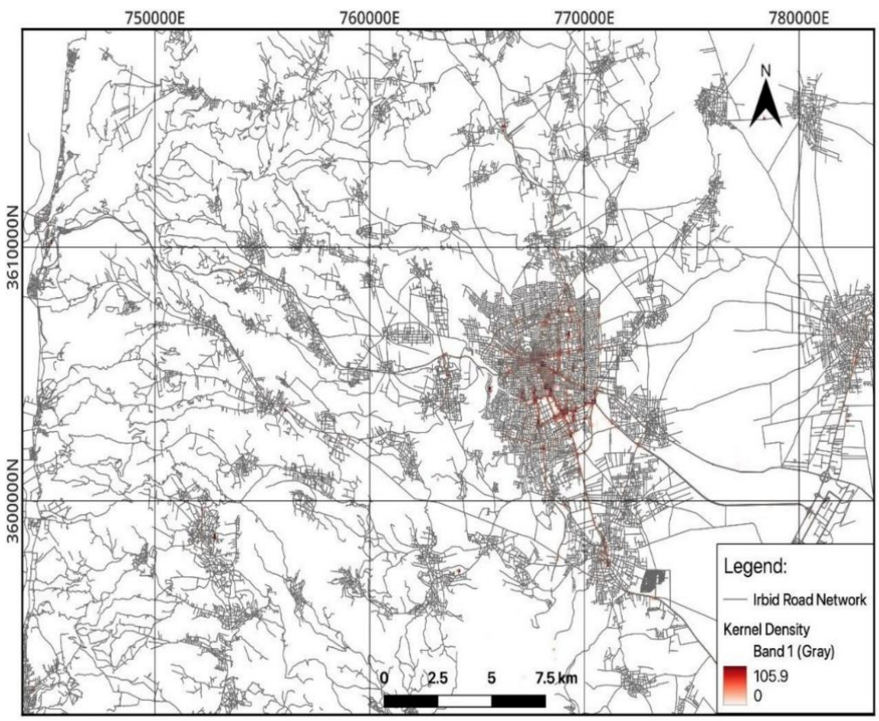

Fig. 5: Kernel density heat map unweighted.

predicted 11,434 instances of low severity accidents with a precision of 0.86, recall of 0.81, and F1-score of 0.84. Similarly, for medium severity accidents, the model achieved a precision of 0.9, recall of 0.93, and F1-score of 0.91, correctly predicting 12,107 instances. In the case of high severity accidents, the model achieved a precision of 0.91, recall of 0.95, and F1-score of 0.93, accurately predicting 10,402 instances. Regarding the "fatalities" severity class, the model accurately predicted 9,645 instances out of the total 10,102 fatality instances. Overall, the weighted average of the precision, recall, and F1-score of the model was 0.9.

3.4 SHAP analysis

In order to explain the predictions and assess the significance of the explanatory variables, the SHAP technique was applied to the final architecture of the accident severity prediction model. Fig. 6 shows the SHAP summary plot, comprehensively explaining how the explanatory variables affect the model's predictions. Each point in the summary plot represents a Shapley value for a feature and an instance. The position of a feature on the y-axis indicates its importance, and the x-axis represents the Shapley value. Additionally, the color of each data point represents the corresponding value. It becomes evident that the speed of the accident exerted the greatest influence on the model's predictions. Subsequently, the accident type, followed by the accident date (year, month, and day of the week), and the primary road class, were identified as factors contributing to the model's predictions. Conversely, road classes such as unclassified, construction, and service had a relatively minimal impact on the model's predictions. This was followed by the link class (primary link), weather conditions, and road surface conditions, in decreasing order of importance.

3.4.1 Impacts of accident-related variables

The analysis revealed a notable association between higher accident speed values and higher severity classes. Accidents occurring at high speeds tend to result in more severe injuries and fatalities, aligning with previous research highlighting the positive correlation between speed and severity class.[42] Additionally, the accident type exhibits a positive correlation with the accident severity class. Collisions and accidents involving a ROR scenario are linked to severe injuries and fatalities, whereas pedestrian accidents tend to be associated with lower severity injuries.

In analyzing the influence of various temporal factors, it is apparent that the day of the week variable exhibits a negative correlation with accident severity class, while the weekend variable demonstrates a positive correlation. This suggests that higher severity classes are more prevalent during weekends and the beginning of the week. Conversely, no clear pattern emerges when considering the accident severity class in relation to the year and month of the year variables. In summary, the findings highlight the significant impact of accident speed and type on the severity of injuries and fatalities. Additionally, the day of the week and weekends play a role in determining the accident severity class.

3.4.2 Impacts of road and weather-related variables

Upon observation, it becomes evident that lower accident severity classes are closely associated with primary, secondary, tertiary, service, and construction road classes. Additionally, primary and tertiary link classes also exhibit a correlation with lower severity classes. In terms of road geometry, segments characterized by being straight, as well as level and upgrade curves, are linked to lower accident severity classes. Conversely, higher severity classes are associated with down grade curves. Higher accident severity classes, on the other hand, demonstrate an association with unpaved road surface types and dry surface conditions. Furthermore, there is a negative correlation between accident severity class and weather conditions, suggesting that lower severity classes tend to occur during adverse weather conditions. This finding can be attributed to drivers exercising increased caution and adopting reduced speeds when confronted with poor surface and weather conditions.

Table 4: Testing confusion matrix and classification report.

Testing dataset					
Actual/Predicted	PDOs	Low severity	Medium severity	High severity	Fatalities
PDOs	13726	1451	365	113	37
Low injury	1279	11434	618	412	308
Intermediate injury	206	204	12107	269	197
High injury	54	76	261	10402	209
Fatalities	31	67	137	222	9645
		Testing classification report			
		Precision	Recall	F1-Score	Support
PDOs		0.9	0.87	0.89	15692
Low injury		0.86	0.81	0.84	14051
Intermediate injury		0.9	0.93	0.91	12983
High injury		0.91	0.95	0.93	11002
Fatalities		0.93	0.95	0.94	10102
		Overall performance metrics			
Accuracy			0.9		63830
Macro avg		0.9	0.9	0.9	63830
Weighted avg		0.9	0.9	0.9	63830

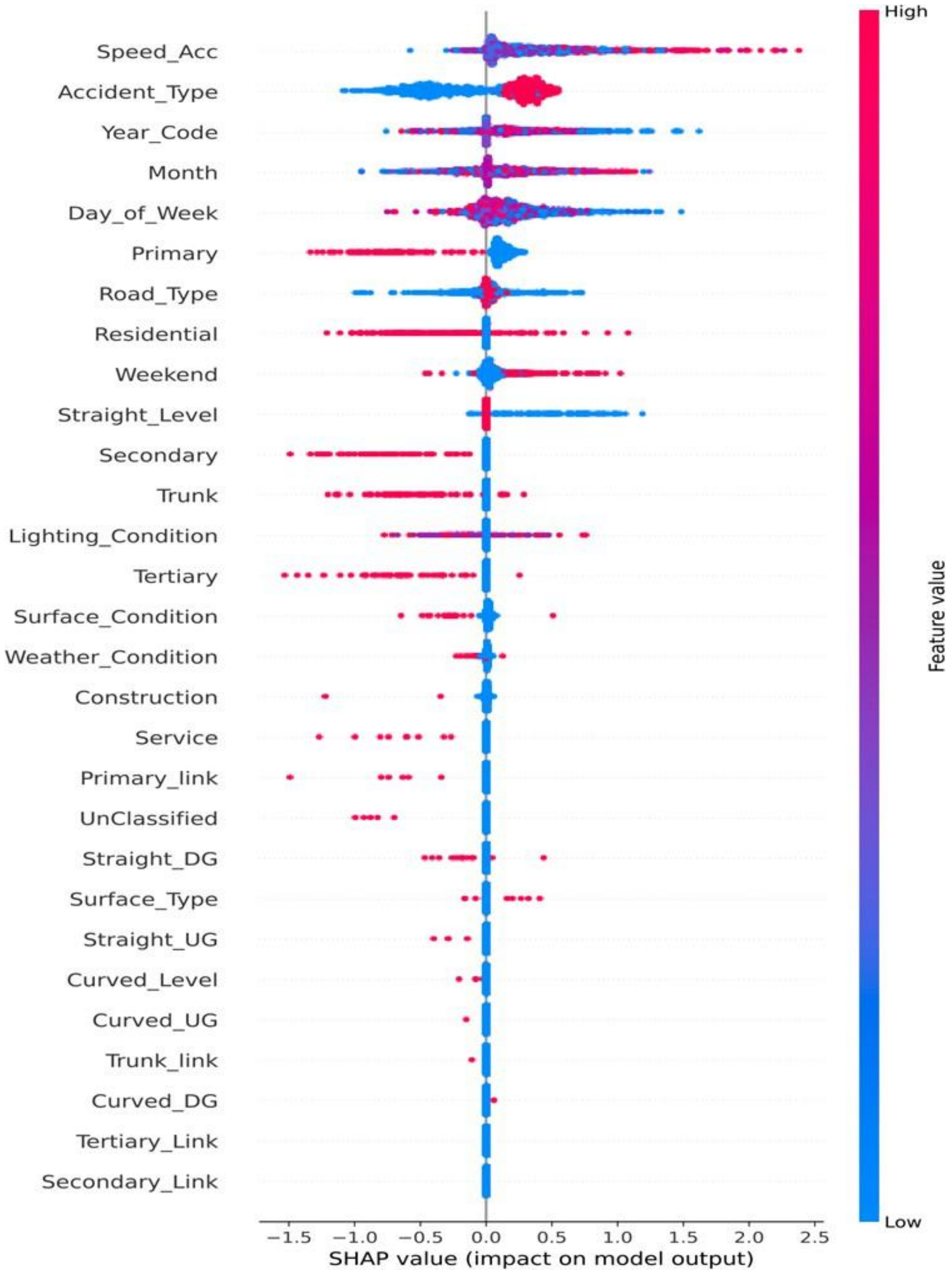

Fig. 6: SHAP analysis.

However, the impact of lighting conditions on accident severity class remains inconclusive and does not exhibit a clear pattern. This study conducted a descriptive analysis of accidents in Irbid from 2016 to 2021, examining both temporal and spatial patterns. Additionally, the study aimed to develop an interpretable data-driven model to predict accident severity. The objectives were to gain a deeper understanding of traffic accident patterns in the city. This was to provide valuable insights for policymakers implementing remedial measures. It was also to offer a severity prediction model for stakeholders such as safety engineers, healthcare institutions, and insurance companies.

The hotspot analysis conducted identified high-risk areas: King Abdulla II sections 2 and 3, Shafiq Irsheidat sections 2 and 3, Prince Hasan sections 1 and 2, and Umar Almokhtar section 1. This aligns with Al-Omari *et al*.'s study on Irbid traffic accidents.[11] However, the study by Al-Omari *et al*.[11] did not identify Alhashmi section 1, King Abdulla II section 1, Rateb Bataineh, and all sections of King Husain as hotspots, due to the differences in the size of the study area between the two studies. Several previous studies highlighted that the RF algorithm is a promising method to model traffic accidents.[20-22] Therefore, RF algorithm was employed in this study to model traffic accident severity levels in Irbid. Moreover, SHAP was used to explain the predictions of the optimal RF specification and to identify the relative importance of the input features.

A notable challenge encountered during model construction was the highly unbalanced nature of the data. The majority of observations corresponded to PDO accidents, while fatality accidents were significantly underrepresented. Consequently, the model performed exceptionally well in predicting PDO accidents but poorly in predicting fatality accidents. To address this issue, SMOTETomek, which considers both oversampling and under sampling techniques, was employed resulting in high overall accuracy for the model. The results highlighted the most influential variables affecting severity predictions. These key factors included accident speed, type, day of the week, road type, and road geometry (particularly straight and level roads). Notably, our findings are consistent with those of Al-Mistarehi *et al*.[22] The influence of certain factors (*e.g*., accident speed, accident type) on the accident severity can be generalized, while others may depend on drivers' experience, behavior, and cultural background.[26-28] Overall, the model can be applied to other Jordanian cities as well as countries with similar driver behavior and experience under varying weather conditions, pavement surface conditions, and road geometric designs. However, the factors identified as most significant in our model may not have the same impact on traffic accidents in other countries, where drivers are more experienced or exhibit different driving behaviors. There are some limitations for this study. First, the prediction model for the accident severity level was developed using only one RF algorithm. The features included in the modeling process were selected based on the available data. Having access to furtherly detailed data on the road network (*e.g.*, speed limits, the number of lanes, traffic control devices, and traffic volume) could enhance the performance of the model and would enable us to investigate the impact of the traffic severity levels.

4. Conclusion

The study of traffic accidents holds vital importance, particularly in developing countries, which experience higher accident rates that significantly impact their already fragile economies. In this study, conducted in the city of Irbid over a six-year period (2016-2021), the primary objectives were to investigate the spatial and temporal patterns of accidents, identify potential hotspots in the road network, and develop interpretable data driven accident severity prediction model.

Hotspot identification was carried out using a kernel density unweighted heatmap, which highlighted certain sections as accident-prone areas. The analysis revealed that these sections are mainly located in the city center along the following streets: Alhashmi, K. Abdulla II, Shafiq Irsheidat, Pr. Hasan, Rateb Bataineh, K. Husain, and Umar Almokhtar. To develop an accident severity prediction model, the study implemented the RF algorithm, which yielded an overall predictive accuracy of 0.9. The model exhibited high performance, with an overall weighted average precision, recall, and F1-score of 0.9. The interpretation of the model using SHAP values highlighted the relative importance of various explanatory variables. Accident speed had the highest impact on the model's predictions, followed by accident type, accident date (year, month, and day of the week), and primary road class.

Furthermore, the study provided insights into the impacts of accident-related variables. Higher accident speeds were associated with higher severity classes, and collision and ROR accidents were linked to severe injuries and fatalities. Moreover, accidents with higher severity classes were more likely to occur on weekends and at the beginning of the week. Regarding road and weather-related variables, the study revealed that tangent segments, level, and upgrade curves were associated with lower accident severity classes. Conversely, downgrade curves were associated with higher severity classes. Additionally, the severity of accidents showed a negative correlation with weather conditions, indicating that poor weather conditions were associated with lower severity classes. In terms of future directions, the study aims to develop count models that predict the number of daily accidents on the link level using various machine learning algorithms. This approach will provide a deeper understanding of accident occurrences in the city of Irbid. Moreover, future studies could consider comparing the performance of other advanced machine learning algorithms for predicting traffic accident severity levels.

Conflict of Interest

There is no conflict of interest.

Supporting Information

Not applicable.

References

[1] H. E. Rosen, I. Bari, N. Paichadze, M. Peden, M. Khayesi, J. Monclús, A. A. Hyder, Global road safety 2010-18: an analysis of global status reports, *Injury*, 2022, **56**, 110266, doi:

10.1016/j.injury.2022.07.030.
[2] B. Ala'a, S. A. Altarawneh, L. Z. A. Ghazaleh, T. M. Jaradat, R. Abujudeh, M. Almbaidin, H. Hasan, Navigating road traffic accidents in Jordan: a retrospective exploration of the health and economic impacts, *Cureus*, 2024, **16**, 1-8, doi: 10.7759/cureus.60497.
[3] B. Edries, A. H. Alomari, Forecasting the fatality rate of traffic accidents in Jordan: Applications of time-series, curve estimation, and multiple linear regression models, *Journal of Engineering Science & Technology Review*, 2022, **15**, 70-77, doi: 10.25103/jestr.156.09.
[4] R. I. Abid, Identifying spatial patterns of road accidents in Madaba City by applying Getis-Ord Gi* spatial statistic, *International Journal of Engineering and Advanced Technology*, 2024, **13**, 1-8, doi: 10.35940/ijeat.C4387.13040424.
[5] K. Hazaymeh, A. Almagbile, A. H. Alomari, Spatiotemporal analysis of traffic accidents hotspots based on geospatial techniques, *ISPRS International Journal of Geo-Information*, 2022, **11**, 260, doi: 10.3390/ijgi11040260.
[6] M. A. Khasawneh, A. A. Al-Omari, B. Ganam, Forecasting traffic accidents in Jordan using regression techniques, *Jordan Journal of Civil Engineering*, 2018, **12**, 570-579.
[7] B. W. Al-Mistarehi, A. H. Alomari, R. Imam, T. K. Alnaasan, Investigating the factors affecting speeding violations in Jordan using phone camera, radar, and machine learning, *Frontiers in Built Environment*, 2022, **8**, 917017., doi: 10.3389/fbuil.2022.917017.
[8] Y. Zhang, T. Liu, Q. Bai, W. Shao, Q. Wang, New systems-based method to conduct analysis of road traffic accidents, *Transportation Research Part F: Traffic Psychology and Behaviour*, 2018, **54**, 96-109, doi: 10.1016/j.trf.2018.01.019.
[9] B. N. Kiran, N. Kumaraswamy, C. Sashidhar, A review of road crash prediction models for developed countries. *American Journal of Traffic and Transportation Engineering*, 2017, **2**, 10-25. doi: 10.11648/j.ajtte.20170202.11.
[10] A. Iranitalab, A. Khattak, Comparison of four statistical and machine learning methods for crash severity prediction, *Accident Analysis & Prevention*, 2017, **108**, 27-36, doi: 10.1016/j.aap.2017.08.008.
[11] A. Al-Omari, N. Shatnawi, T. Khedaywi, T. Miqdady, Prediction of traffic accidents hot spots using fuzzy logic and GIS, *Applied Geomatics*, 2020, **12**, 149-161, doi: 10.1007/s12518-019-00290-7.
[12] H. Al-Masaeid, Traffic accidents in Jordan, *Jordan Journa of Civil Engineering*, 2009, **3**, 331-343.
[13] K. Ivan, I. Haidu, The spatio-temporal distribution of road accidents in Cluj-Napoca, *Geographia Technica*, 2012, **2**, 32-38.
[14] M. Bíl, R. Andrášik, Z. Janoška, Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation, *Accident Analysis & Prevention*, 2013, **55**, 265-273, doi: 10.1016/j.aap.2013.03.003.
[15] Z. Xie, J. Yan, Kernel Density Estimation of traffic accidents in a network space, *Computers, Environment and Urban Systems*, 2008, **32**, 396-406, doi: 10.1016/j.compenvurbsys.2008.05.001.
[16] V. Prasannakumar, H. Vijith, R. Charutha, N. Geetha, Spatio-temporal clustering of road accidents: GIS based analysis and assessment, *Procedia - Social and Behavioral Sciences*, 2011, **21**, 317-325, doi: 10.1016/j.sbspro.2011.07.020.
[17] C. Dong, C. Shao, J. Li, Z. Xiong, An improved deep learning model for traffic crash prediction, *Journal of Advanced Transportation*, 2018, **2018**, 3869106, doi: 10.1155/2018/3869106.
[18] S. Malik, H. El Sayed, M. A. Khan, M. J. Khan, Road accident severity prediction: a comparative analysis of machine learning algorithms, 2021 IEEE Global Conference on Artificial Intelligence and Internet of Things, December 12-16, Dubai, United Arab Emirates, IEEE, 2021, 69-74, doi: 10.1109/gcaiot53516.2021.9693055.
[19] S. Ahmed, M. A. Hossain, M. M. I. Bhuiyan, S. K. Ray, A comparative study of machine learning algorithms to predict road accident severity, 20th International Conference on Ubiquitous Computing and Communications, December 20-22, London, United Kingdom, IEEE, 2021, 390-397, doi: 10.1109/iucc-cit-dsci-smartcns55181.2021.00069.
[20] S. S. Yassin, Pooja, Road accident prediction and model interpretation using a hybrid K-means and random forest algorithm approach, *SN Applied Sciences*, 2020, **2**, 1576, doi: 10.1007/s42452-020-3125-1.
[21] B. Geyik, M. Kara, Severity prediction with machine learning methods, International Congress on Human-Computer Interaction, Optimization and Robotic Applications, June 26-28, Ankara, Turkey, IEEE, 2020, 1-7, doi: 10.1109/hora49412.2020.9152601.
[22] B. W. Al-Mistarehi, A. H. Alomari, R. Imam, M. Mashaqba, Using machine learning models to forecast severity level of traffic crashes by R studio and ArcGIS, *Frontiers in Built Environment*, 2022, **8**, 860805, doi: 10.3389/fbuil.2022.860805.
[23] J. Wu, X. Y. Chen, H. Zhang, L. D. Xiong, H. Lei, S. H. Deng, Hyperparameter optimization for machine learning models based on Bayesian optimization, *Journal of Electronic Science and Technology*, 2019, **17**, 26-40, doi: 10.11989/JEST.1674-862X.80904120.
[24] R. Turner, D. Eriksson, M. McCourt, J. Kiili, E. Laaksonen, Z. Xu, I. Guyon, Bayesian optimization is superior to random search for machine learning, *Proceedings of the NeurIPS 2020 Competition and Demonstration Track*, 2021, **133**, 3-26.
[25] K. Jadaan, I. Al-Hyari, H. Naghawi, R. Ammourah, Z. Al Nabulsi, Traffic safety in Jordan: magnitude, cost and potential countermeasures, *Journal of Traffic and Logistics Engineering*, 2013, **1**, 54-57, doi: 10.12720/jtle.1.1.54-57.
[26] T. Özkan, T. Lajunen, J. E. Chliaoutakis, D. Parker, H. Summala, Cross-cultural differences in driving behaviours: a comparison of six countries, *Transportation Research Part F: Traffic Psychology and Behaviour*, 2006, **9**, 227-242, doi: 10.1016/j.trf.2006.01.002.
[27] H. Abuzaid, R. Almashhour, G. Abu-Lebdeh, Driving towards sustainability: a neural network-based prediction of the traffic-related effects on road users in the UAE, *Sustainability*, 2024, **16**, 1092, doi: 10.3390/su16031092.
[28] O. A. Taiwo, S. A. Hassan, R. B. Mohsin, N. Mahmud,

Analysis of accident predictability and the use of driver behaviour questionnaire: a systematic review, *International Journal of Research and Innovation in Social Science*, 2024, **8**, 2347-2366, doi: 10.47772/ijriss.2024.803164.
[29] T. Khedaywi, N. AlKofahi, L. Alomarri, Development of a statistical model for predicting a traffic noise (case study: Irbid-Jordan), *Jordan Journal of Civil Engineering*, 2021, **15**, 551-561, doi: 10.2174/1874447801307010001.
[30] S. Al-kheder, N. Haddad, L. Fakhoury, S. Baqaen, A GIS analysis of the impact of modern practices and polices on the urban heritage of Irbid, Jordan, *Cities*, 2009, **26**, 81-92, doi: 10.1016/j.cities.2008.12.003.
[31] M. M. Awawdeh, R. R. Abuhadba, M. M. Jamhawi, A. I. Rawashdeh, R. N. Jawarneh, M. M. Awawdeh, Urban expansion in Greater Irbid Municipality, Jordan: the spatial patterns and the driving factors, *GeoJournal*, 2024, **89**, 43, doi: 10.1007/s10708-024-11036-3.
[32] B. W. Silverman, Density estimation for statistics and data analysis, New York, Routledge, 1998, ISBN: 9781315140919.
[33] L. Breiman, Random forests, *Machine learning*, 2001, **45**, 5-32, doi: 10.1023/A:1010933404324.
[34] S. M. Lundberg, S. I. Lee, A unified approach to interpreting model predictions, *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017, **30**, 4768-4777, doi: 10.48550/arXiv.1705.07874.
[35] T. K. Anderson, Kernel density estimation and K-means clustering to profile road accident hot spots. *Accident Analysis & Prevention*, 2009, **41**, 359-364, doi: 10.1016/j.aap.2008.12.014.
[36] L. Yang, A. Shami, On hyperparameter optimization of machine learning algorithms: Theory and practice, *Neurocomputing*, 2020, **415**, 295-316, doi: 10.1016/j.neucom.2020.07.061.
[37] N. Alsaleh, B. Farooq, The impact of COVID-19 pandemic on ridesourcing services differed between small towns and large cities. *PLoS ONE*, 2022, **17**, 1-22, doi: 10.1371/journal.pone.0275714.
[38] N. Alsaleh, & B. Farooq, Interpretable data-driven demand modelling for on-demand transit services, *Transportation Research Part A: Policy and Practice*, 2021, **154**, 1-22, doi: 10.1016/j.tra.2021.10.001.
[39] A. Kalatian, B. Farooq, Decoding pedestrian and automated vehicle interactions using immersive virtual reality and interpretable deep learning, *Transportation Research Part C: Emerging Technologies*, 2021, **124**, 1-25, doi: 10.1016/j.trc.2020.102962.
[40] I. Twick, G. Zahavi, H. Benvenisti, R. Rubinstein, M. S. Woods, H. Berkenstadt, A. Nissan, E. Hosgor, D. Assaf, Towards interpretable, medically grounded, EMR-based risk prediction models, *Scientific Reports*, 2022, **12**, 1-12, doi: 10.1038/s41598-022-13504-7.
[41] J. Lever, M. Krzywinski, N. Altman, Points of Significance: Model selection and overfitting. *Nature Methods*, 2016, **13**, 703-704, doi: 10.1038/nmeth.3968.
[42] R. Elvik, Speed and road safety: synthesis of evidence from evaluation studies, *Transportation Research Record*, 2005, **1908**, 59-69, doi: 10.1177/0361198105190800108.