Severity Prediction of Road Accidents in Jordan using Artificial Intelligence

Dheya Mustafa[1,*], Mohammad F. Al-hammouri[2], Safaa M. Khabour[3]

[1,2]Department of Computer Engineering, Faculty of Engineering, The Hashemite University, Zarqa 13133, Jordan

[3]Department of Information Systems, Yarmouk University, Irbid 21163, Jordan

(Received: December 27, 2024; Revised: January 5, 2025; Accepted: March 15, 2025; Available online: April 15, 2025)

Abstract

Road traffic accidents are a significant global concern, with developing countries accounting for 85% of annual fatalities and 90% of disability-adjusted life years lost. This study investigates the severity of road accidents in Jordan using a machine learning-based predictive approach. A dataset of 73,000+ accident reports from 2018 was analyzed, covering factors such as road conditions, weather, vehicle attributes, and driver demographics. The primary objective is to develop and evaluate machine learning models for predicting accident severity. Seven classification algorithms were tested: Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naïve Bayes (NB), Random Forest (RF), Decision Tree (DT), and Extreme Gradient Boosting (XGBoost). The results indicate that LR achieved the highest accuracy at 98.1%, followed by RF (95.02%) and XGBoost (95.27%). Feature importance analysis revealed that road type, lighting conditions, and driver violations were the most influential factors in predicting accident severity. A key novelty of this research is the integration of real-world Jordanian accident data with machine learning models to enhance predictive accuracy. The study's findings provide actionable insights for policymakers, enabling targeted interventions to reduce accident severity. The dataset is made publicly available to support future research. This research contributes to the advancement of AI-driven traffic safety solutions, demonstrating the effectiveness of machine learning in real-time risk assessment and decision-making.

Keywords: Accident Severity Prediction, Car Accident, Collision Fatalities, Machine Learning

1. Introduction

There are a lot more auto accidents these days, and the costs incurred from traffic-related fatalities and injuries have a big influence on society [1], [2]. Several research efforts aim to pinpoint the key variables that majorly impact the severity of injuries from auto accidents. Various variables present at the scene, such as the weather, road conditions, driver age, automobile type and other elements that may influence traffic accident rates, can provide a fair estimate of the likelihood of an accident. Auto accidents cause many fatalities and large financial losses each year.

Jordan is a Middle Eastern nation comprising 90,000 square kilometers and home to about 11 million people. Road accidents escalated to a serious issue in the middle of the 1980s [3]. In Jordan, automobile accidents were the second leading cause of death in 2007. Accidents increased from 15,884 in 1987 to 110,630 in 2007 during the last 20 years. The population and the quantity of cars grew by almost two and three times, respectively, during that time. Jordan's traffic accident rate is high and still rising; in 2021, there were 160,600 incidents. Of those, 11,241 resulted in injuries to people, 589 in fatalities, and 320 million JOD losses. Currently ranking eighth globally in terms of cause of death, traffic accidents claim the lives of 1.35 million people each year. Governments worldwide must comprehend the main reasons behind these incidents and the surrounding conditions before implementing policies to reduce traffic-related fatalities. This situation forced the administration to assemble an extensive plan to deal with the problem [4], [5], [6].

A substantial amount of work has been devoted to studying real-world traffic crash data to gain a better understanding of road traffic operating challenges and to identify hazardous road portions. Identifying these risk factors aids in the development, execution, and evaluation of highway safety programs that tend to minimize road traffic crashes and

***Corresponding author: Dheya Mustafa (dheya@hu.edu.jo)**
DOI: https://doi.org/10.47738/jads.v6i2.624
This is an open access article under the CC-BY license (https://creativecommons.org/licenses/by/4.0/).
© Authors retain all copyrights

improve vehicle safety features. Furthermore, the findings of such research help decision-makers build long-term, statewide strategic plans for traffic and highway safety, improve post-accident treatment for road collision victims and help highway safety administrators raise public awareness.

A road accident is generally described as an incident on a public or private road that occurs due to someone's negligence or omission or by an external element (a natural catastrophe excluded), culminating in a collision. This includes uncontrollable incidents and collisions where victims in a car collide with something either inside or outside the vehicle, like a passenger on a bus.

There is at least one moving vehicle engaged, and the police have documentation of any damage or injuries to people, property, vehicles, buildings, or animals. Most often, several complex factors interact to result in traffic accidents. According to research studies, we may infer the key factors affecting traffic accidents. The interactions of the car, the driver, the road, and the surroundings cause traffic accidents. Numerous causes can lead to accidents, but the most frequent ones are those linked to the driver, the road, the surroundings, and the vehicle.

This work extends the previous preliminary project [7]. It aims to gain knowledge on the severity of traffic accidents in Jordan currently, as well as the effectiveness of several machine-learning approaches in predicting the mortality rates of those accidents [8], [9], [10]. To forecast the severity of traffic accidents in Jordan, this study describes and investigates the development of a prediction model utilizing data science and machine learning techniques [10]. The research effort is divided into two stages. First, the data needs to be cleaned and preprocessed to be ready for Phase 2. Then, the dataset will be used to train and test appropriate algorithms using Python [11], such as Naïve Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Extreme Gradient Boosting (XGBoost) and others [12], [13].

This study explores critical research questions related to traffic safety in Jordan. It seeks to determine the percentage of severe accidents occurring in the country, providing a clearer understanding of the extent and gravity of the issue. Based on its findings, the study aims to offer insightful recommendations that could contribute to improving traffic conditions and enhancing road safety. Additionally, it examines the effectiveness of machine learning in predicting the severity of collisions on Jordanian roads, assessing its accuracy and potential role in mitigating accident risks through data-driven insights. To answer these questions, we obtained real data from Jordanian authorities on road incidents that occurred in 2018. The data was then preprocessed and thoroughly analyzed. Data was made available online to academics. Many experiments were carried out on the preprocessed data using various machine-learning techniques to implement and evaluate the accident severity prediction method.

This work contributes to the field by collecting, translating, and analyzing real traffic accident data from Jordan, making it accessible online for future research. Additionally, it develops, tests, and evaluates a machine learning approach for predicting accident severity, providing valuable insights that can aid in traffic management and accident prevention.

The remainder of this paper is organized as follows: The following section critically reviews related works. Section 3 describes the research methodology. Section 4 discusses evaluation experiments and their results. Finally, in Section 5, we wrap up our findings.

2. Related Work

Several road accident analysis researchers have attempted to develop prediction models that automatically classify different traffic accidents according to the severity of injuries sustained and pinpoint the patterns associated with high-risk situations [14]. Related studies are compared and summarized in table 1.

Chong et al. [15] proposed models that use three machine-learning techniques—hybrid models of decision trees and neural networks (DTANN), hybrid neural networks (NN), and decision trees (DT)—to predict the severity of injuries incurred in auto accidents. Their algorithms attempted to categorize the injury kind into five groups: no injury, potential injury, non-incapacitating injury, incapacitating injury and fatal injury. The results showed that employing DTANN, the models for fatal and non-fatal injuries performed better than those in other classes, reaching 90% and 72.63%, respectively.

Table 1. Literature review summary

Source	Techniques	Accident Categories	Best Accuracy	Dataset Country	Num. of Factors	Steps to resolve Challenges in the Dataset
(Chong et al., 2004) [15]	SVM, DT, DTANN, NN	1. No injury. 2. Possible. 3. Non-incapacitating. 4. Incapacitating. 5. Fatal injury.	90% for fatal injury using DTANN.	United States.	25	Dataset cleaning and handling unknown values.
(Atwah and Al-Mousa, 2021) [16]	ANN, RF, SVM, hard voting ensemble	1. Slight. 2. Serious. 3. Fatal.	83.9% using RF.	United Kingdom.	27	Dataset cleaning, balancing classes, one-hot encoding, features scaling (normalization and standardization)
(Patel et al., 2021) [17]	RF	3 levels of casualty severity	87% using RF.	United Kingdom.	31	Redundancy and null elimination, feature scaling, and features reduction.
(Yassin et al., 2020) [18]	SVM, K-means, K-means-RF, RF, KNN, LR, ANN	1. Fatal 2. Severe. 3. Light.	99.86% using K-means-RF.	Ethiopia.	14	Data cleaning, missing value handling, outlier treatment, dealing with absolute value, encoding and normalization.
(Bokaba et al., 2022) [19]	SVM, LR, RF, NB, AdaBoost, kNN	1. Major. 2. Minor. 3. Natural disaster. 4. None. 5. Unknown.	97% using RF.	Gauteng province.	8	Dataset cleaning and handling missing values, PCA and LDA dimensionality reduction and preparation.
(Hashmienejad and Hasheminejad, 2017) [20]	SVM, NB, DT, Rule based, KNN, ANN, Customized (NSGA-II)	1. Accidents' occupants are injured slightly. 2. Accidents' occupants are killed.	88.2% using the customized (NSGA-II).	Iran.	25	Replacing missing and removing noise.
(Yan and Shen, 2022) [21]	ANN, KNN, SVM, RF, BO-RF	1. Slight. 2. Serious. 3. Fatal.	96.25% using BO-RF.	United States.	15	Missing values, filling variables, and coding variables.
(Yousif and AlRababaa, 2013) [22]	MLP	Classify accident types: Collision, pedestrian, or overturn	100% in identifying accident type	Jordan	4	Dataset preprocessing, normalization and encoding.
(Sameen, M. I., and Pradhan, B. 2017) [23]	RNN	Severity classes: -disabling/fatality -possible/evident injury -property damage only	71.7% using RNN	Malaysia	9	Removing missing data, one hot encoding, and detection of highly correlated factors. Mitigating overfitting using Gaussian noise injection into the training data, using a ReLU activation function in the hidden layers, and applying the dropout technique.
Geyik, B., and Kara, M, 2020) [24]	MLP	1. Fatal 2. Serious 3. Slight	86.67% using MLP	United Kingdom.	17	Data reduction, feature sampling, transformation, and checking the missing values.

Promising results are obtained when utilizing Random Forest (RF) to classify the severity of traffic accidents. SVM, hard voting ensemble model, RF, and ANN were the four classifiers that Atwah and Al-Mousa [16] worked with. There are three categories for this kind of injury: minor, major, and fatal. The RF model offered the highest degree of accuracy, coming in at 83.9%. Patel et al. [17] reported an 87% accuracy for the RF model. They recommended the best insurance plan based on a detailed analysis of auto accidents, categorized by age groups and weekdays. Yassin et al. [18] presented a hybrid approach combining K-means and RF for accident severity categorization. K-means leveraged the ability to extract hidden information from traffic accident data to create a new feature in the training set. Severity was categorized and predicted using RF, outperforming other traditional models such as SVM, Logistic Regression (LR) and KNN, with an overall accuracy of 99.86%. Furthermore, Bokaba et al. found that the RF classifier combined with multiple imputations through chained equations yielded the best classification performance [19]. The association between the variables related to road traffic accidents (RTAs) was ascertained using principal component analysis (PCA) and linear discriminant analysis (LDA), and the model's performance was enhanced through the application of dimensionality reduction.

Based on user preferences, Hashmienejad and Hasheminejad introduced a novel rule-based method for predicting the severity of traffic accidents [20]. The Non-Dominated Sorting Genetic Algorithm (NSGA-II), a multi-objective genetic algorithm, was modified in the proposed method to optimize and identify rules that adhere to the Support, Confidence, and Comprehensibility metrics. The accuracy of the model is 88.2%. Yan and Shen [21] proposed a hybrid model that blends RF with Bayesian optimization (BO). In the proposed model (BO-RF), RF is the base predictive model, while BO is utilized to optimize RF's parameters.

Yousif and AlRababaa [22] proposed a Multi-Layer Perceptron (MLP) model for predicting traffic accidents in Jordan using a neural network. They used a comprehensive dataset, including features such as accident types, causes, and environmental driving conditions. The MLP neural network showed a high accuracy, achieving approximately 100% accuracy in predicting accident types. Sameen and Pradhan [23] utilized a Recurrent Neural Network (RNN) model to classify accident severity into three categories, achieving 71.7% accuracy. Geyik and Kara [24] employed an MLP model to classify accident severity into fatal, serious, and slight categories, achieving an accuracy of 86.67%. AlKofahi [25] studied the effects of refugees on traffic accident trends in Jordan, analyzing accident data from 1981 to 2019. The results indicate that the presence of refugees has a minimal impact on traffic accident development. However, the increase in the number of vehicles had a more significant influence.

The application of machine learning techniques to model traffic accident severity has gained significant attention in recent years [14], [21], [26], [27], [28]. Traditional statistical methods, such as LR and DT, have been supplemented by more advanced machine learning approaches that offer improved accuracy and predictive power. SVM is effective for high-dimensional data but struggles with imbalanced datasets. RF and XGBoost are robust for complex interactions but may overfit irrelevant features. LR excels in interpretability but may struggle with complex interactions without feature engineering.

Recent studies have explored a variety of machine learning models to predict accident severity, reflecting the diversity of approaches in this field. For instance, Tayebi et al. employed a Gradient Boosting Machine (GBM) to analyze traffic accidents in Canada, demonstrating that boosting techniques can significantly enhance prediction accuracy by effectively capturing complex patterns in the data [28]. This study highlighted the importance of using ensemble methods to improve model robustness and accuracy.

Incorporating hybrid methods that combine multiple algorithms is becoming increasingly popular [29]. Sun et al. introduced a hybrid model combining RF and Neural Networks to predict accident severity in China . Their approach leveraged the strengths of RF for feature selection and Neural Networks for capturing non-linear relationships, resulting in a model that outperformed individual algorithms.

Another notable study by Yang et al. developed a hybrid ensemble model combining XGBoost with a Bayesian Network to predict traffic crash severity [30]. By analyzing the interaction between road and environmental factors, the model achieved a prediction accuracy of 89.05%. Their findings highlight the importance of considering multi-factor interactions in crash severity prediction and demonstrate the effectiveness of integrating machine learning with probabilistic models for improved predictive performance.

Ensemble learning techniques, such as bagging and boosting, have shown significant promise in improving the accuracy of accident severity models. Zhao et al. utilized the AdaBoost algorithm to enhance the predictive performance of DT models, resulting in a more accurate classification of accident severity in urban areas [31]. This study underscores the benefits of boosting algorithms in reducing variance and improving model stability.

3. Methodology

This study's main goal was to assess and contrast the predictive abilities of seven regression algorithms—SVM, KNN, NB, LR, XGBoost, RF and DT—to determine how well they could forecast the severity of accidents. The research methodology is illustrated in figure 1.

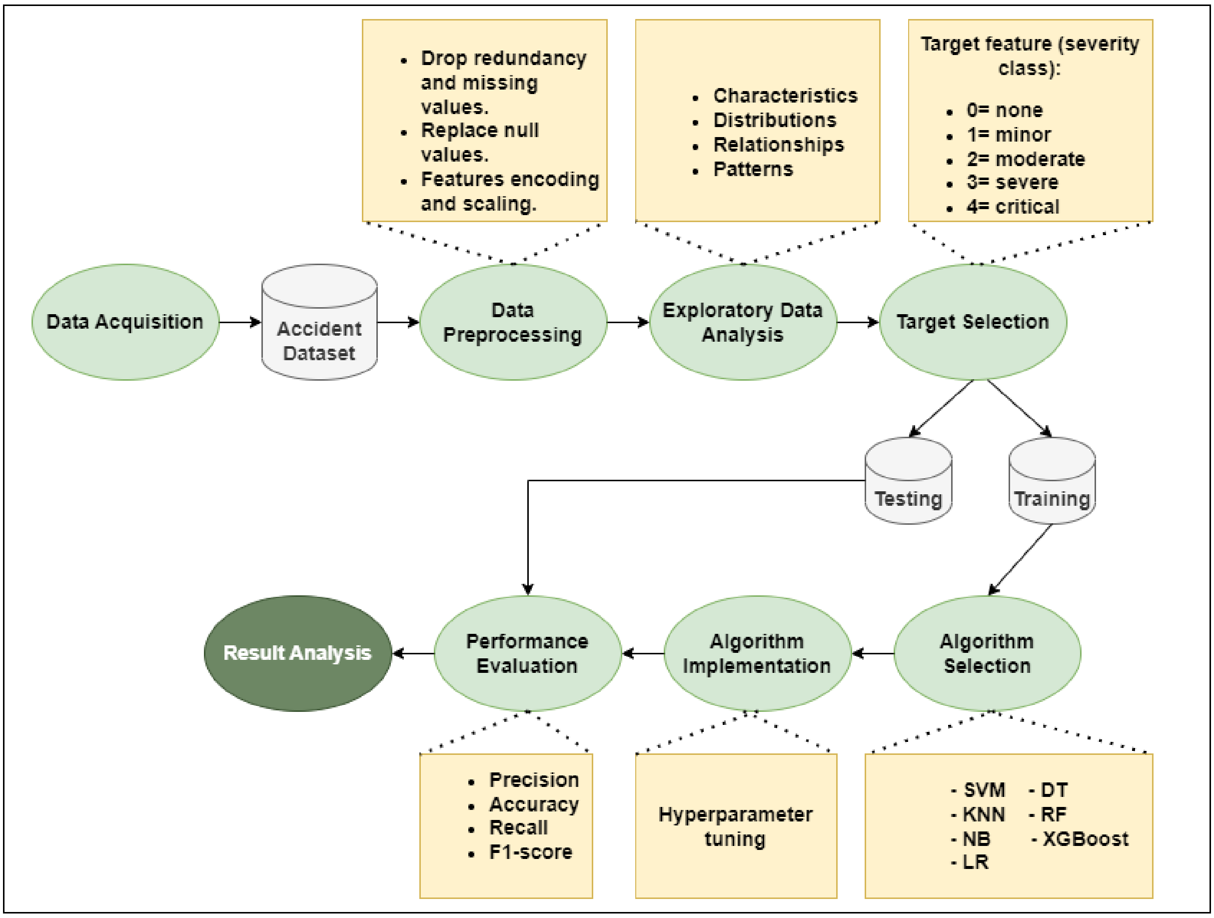

Figure 1. The research methodology.

The design of this study follows a structured approach, beginning with data acquisition, where a dataset comprising records of 73,000 actual traffic incidents on Jordanian roads was obtained from the Jordanian Traffic Department. This dataset includes comprehensive details about the conditions surrounding each accident. The next phase involves data preprocessing, which entails gathering, cleaning, and preparing the data to ensure its suitability for training and evaluating regression algorithms. Outliers and missing values are removed, and necessary feature engineering is performed to refine the dataset based on its actual characteristics.

To gain deeper insights into the dataset, exploratory data analysis is conducted to examine properties, distributions, and relationships between variables. This step helps identify patterns and trends that could influence the performance of regression algorithms. Target selection follows, where a linear transformation model is used to create a derived feature called severity rate, a normalized weighted average based on the quantity and types of injuries. This method was chosen for its simplicity in aggregating and scaling variables while maintaining interpretability and capturing meaningful relationships in a linear manner. While effective in this context, the study acknowledges that more advanced techniques, such as PCA or tree-based feature engineering, could uncover more complex relationships, and future work will explore these alternatives.

The algorithm selection phase involves choosing seven classification algorithms—SVM, KNN, NB, LR, XGBoost, RF, and DT—based on their suitability for the dataset and their established strengths and weaknesses. Once selected, these algorithms are implemented with careful tuning of hyperparameters to optimize their predictive performance. Performance evaluation is then conducted by running the trained algorithms on the dataset and assessing their accuracy in predicting accident severity. Finally, the analysis of results involves comparing the performance of these models to determine which algorithm demonstrates the highest predictive accuracy.

This study, by fulfilling these project and design objectives, attempted to offer important insights into the efficacy of various regression algorithms for predictive modelling in the provided dataset. The research findings can potentially influence future applications in related sectors and guide decision-making.

3.1. Dataset Description

The JO-Traffic-Accidents-Dataset (JO-TAD) was used for this study, which includes information on traffic accidents in Jordan. The Central Traffic Department, Public Security Directorate, Amman, Jordan, provided the raw dataset in Arabic. JO-TAD includes 73,095 accident reports from 2018, emphasizing accidents that happened in January, June, and September. Winter, summer, and fall are the three separate seasons that these months correlate to.

When working on a dataset pertaining to the forecast of accident severity, taking into consideration temporal changes such as the increase in the number of vehicles, improvements in road infrastructure, or modifications to traffic legislation can greatly improve the study. We address this by dividing the dataset into quarterly intervals so that we may track changes in accident trends over time. The seasons of winter, summer, and fall are represented by each quarter. This aids in spotting patterns or irregularities.

The Jordanian Road Traffic Accidents Reports Dataset is publicly available as an Excel file (JO_traffic_accidents_dataset.xlsx) and a CSV file (JO_traffic_accidents_raw_data_En.csv) at the Mendeley website (Link: https://doi.org/10.17632/r6db558376.1). JO-TAD offers various features, with its main categories, visualizations, and statistics detailed in the following subsections.

3.1.1. Accident Statistics

Accident types were categorized into three main groups: Off-Road, Run-over, and Collision. Table 2 presents the classification of accident types alongside their corresponding frequencies. The data show that collisions represent the predominant type of accident. Figure 2 shows the distribution of accidents by speed, while figure 3 shows the accident distribution in Jordan cities. Amman, Jordan's capital and largest city, is the most accident-prone due to its high population and car density, leading to more clogged roads and increased collision risks. Figure 4 illustrates the distribution of accidents throughout three quarters of the year 2018: January, June, and September, respectively.

Winter, specifically the first quarter (January), is the most frequent season for accidents due to factors such as snow, ice, and sleet limiting tire traction, poor vision due to fog, heavy snow, and fewer daylight hours, and frost or condensation on windshields. Wet or slick roads also lengthen stopping distances, making it difficult to dodge obstructions or other cars' abrupt stops. Cold conditions also impair car performance, increasing the risk of collisions.

Table 2. Classification of Accident Type

Accident types	Frequency	Percent
Off-Road	463	0.63
Run over	874	1.19
Collision	71758	98.17

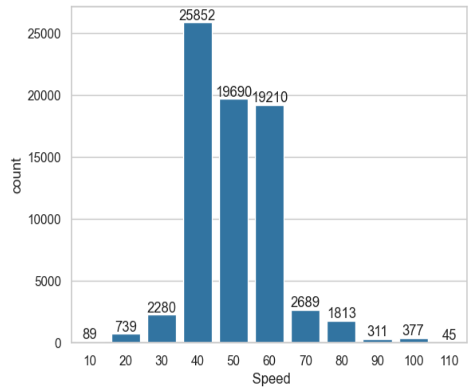

Figure 2. The distribution of accidents by Speed.

Fall season in the third quarter (September), marked by rain, fog, and wet leaves, increases road accidents due to weather changes, road conditions, and human behavior. Shorter days and earlier sunsets can hinder driver visibility, disrupt sleep patterns, and increase the risk of accidents. Back to school activities also increase the risk of accidents. Temperature variability can cause fog and condensation on windshields and roadways, making it harder to see.

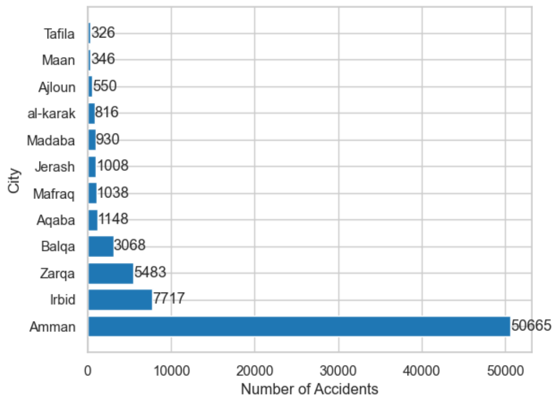

Figure 3. Accident Distribution in Jordan Cities

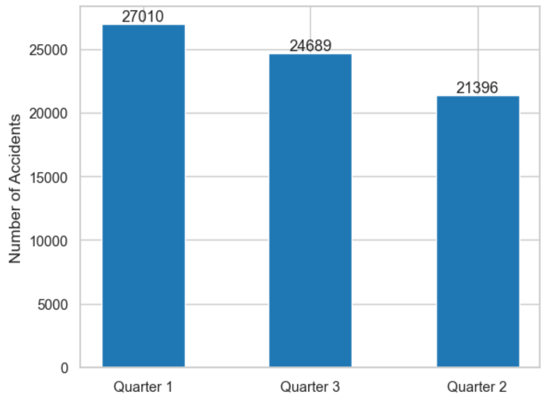

Figure 4. Distribution of the number of accidents in the three quarters of the year 2018.

3.1.2. Driver Information

The JO-TAD dataset contains features related to the drivers involved in accidents, excluding passengers. Specifically, we focus on driver attributes such as sex, age, license type, and Driver mistake. For example, due to health status, the possibility of an accident increases with age [32]. Figure 5 illustrates the distribution of driver sex classifications. The analysis indicates that most accidents, comprising 77%, involve male drivers, while female drivers are involved in 10.66%. Additionally, 12.33% of accidents are attributed to unknown drivers, which could be due to either the driver escaping the scene or the police officer failing to record the driver's information accurately. Figure 6 shows the distribution of Driver Age using the histogram.

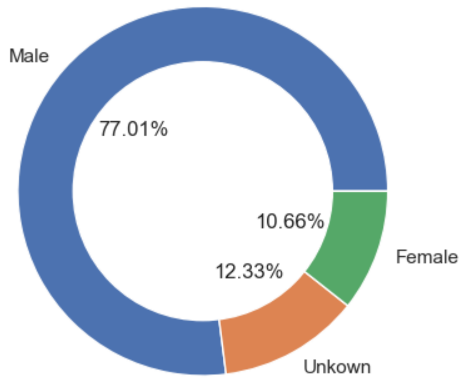

Figure 5. Classification of driver sex.

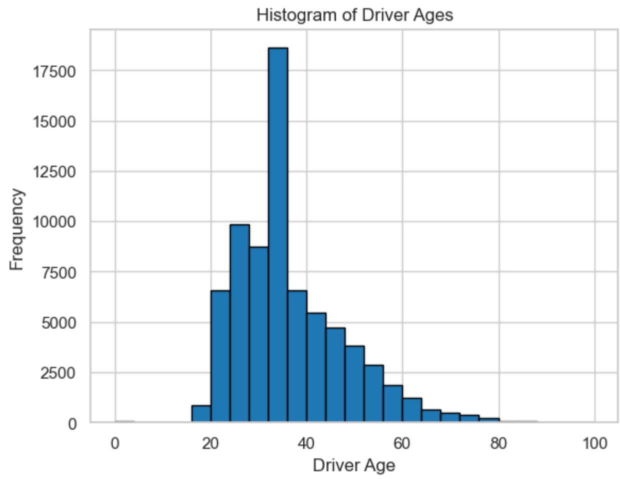

Figure 6. The distribution of Driver Age using histogram.

One important feature that should be addressed is the involvement of driver mistakes in accidents. We can effectively relieve or even eradicate many accidents by carefully examining and categorizing these mistakes. Figure 7 illustrates the distribution of driver mistakes within the JO-TAD dataset. Specifically, it highlights the top 15 driver mistakes associated with accidents.

Notably, "violation of traffic rules" emerges as the most common mistake, accounting for 40,962 occurrences. This is followed by tailgating, where drivers fail to maintain a safe distance. It's worth mentioning that although the driver mistake feature comprises 120 different categories, similar mistakes were grouped to summarize the number of accidents related to each category, as illustrated in figure 7. The Dataset Description file provides a comprehensive summary of all driver mistakes.

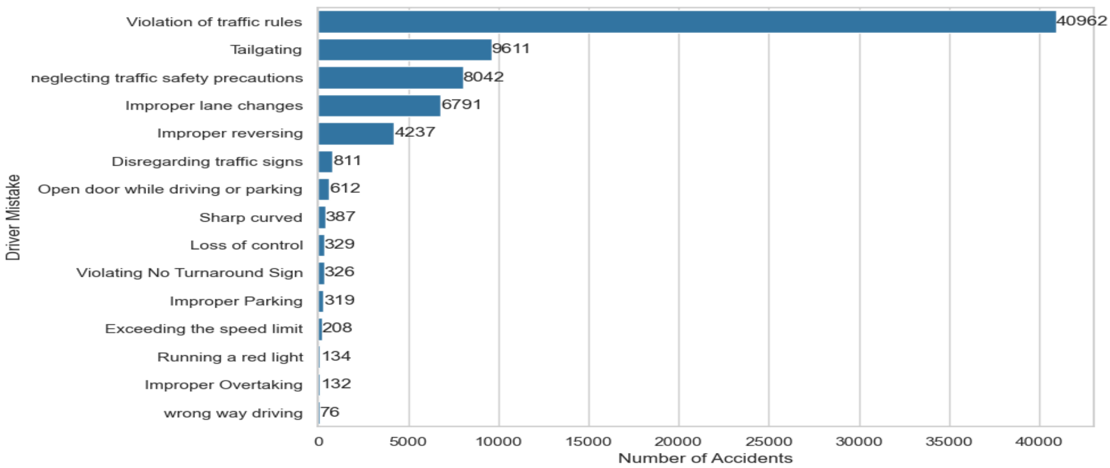

Figure 7. Count of Accidents by Top 15 Driver Mistakes.

The final feature to discuss within this context is the driver's license type, which comprises 12 categories. The most common license type is class, accounting for 47,622 instances. This type enables drivers to operate small passenger vehicles (up to 9 seats), which is understandable. The categories and their corresponding counts are detailed within the data description file.

3.1.3. Accident Severity

The primary features associated with accident severity and injury types are categorized from lowest to highest severity: Simple Injuries, Medium Injuries, Severe Injuries, and Death. Figure 8 illustrates a summary count of the various severity types across the three quarters. The figure highlights that most accident injuries are classified as simple injuries, while the percentage of fatalities is comparatively lower than other severity types.

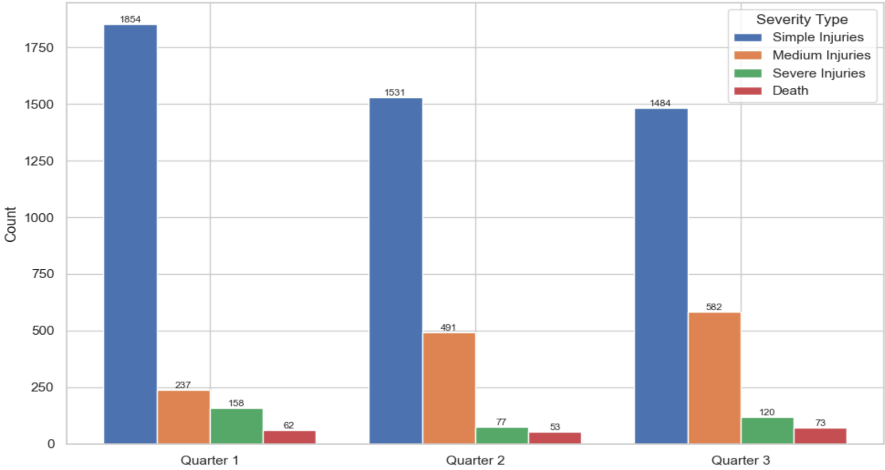

Figure 8. Summary of Various Severity Types by Quarter.

Figure 9 shows the distribution of accident severity by accident type. The data indicates that collisions demonstrate the highest severity rate, which aligns with expectations, as collisions constitute the most prevalent type of accidents.

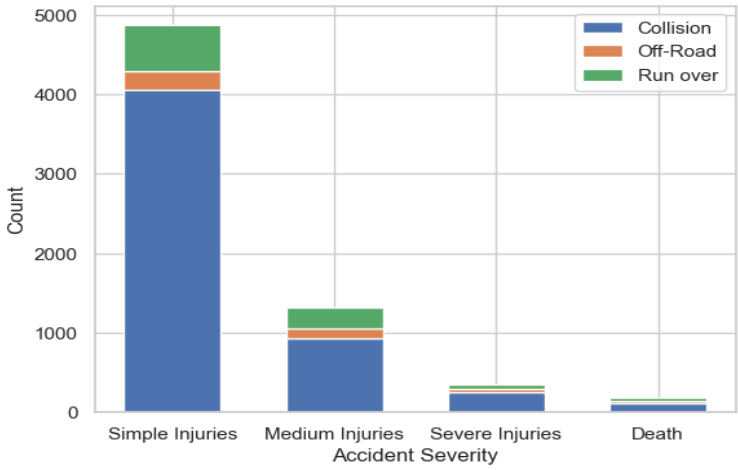

Figure 9. Distribution of Accident Severity by Accident Type.

3.1.4. Driving Conditions

The JO-TAD dataset includes attributes related to driving conditions, including factors such as daylight and weather, such as Light and Weather. Figure 10 shows that the majority, 70.2% (51285 accidents), occur during daylight with sufficient light. This is followed by 20.6% of accidents occurring at night with enough light. The categories and their respective counts are presented in the dataset description file for the weather attribute. It's worth noting that most accidents occur in clear weather conditions.

Other driving conditions are road-related factors such as Road Lanes, Road Type, Road Surface Description, and Road Properties. Figure 11 illustrates the distribution of Road Lanes categories. The majority, accounting for 56.2% (41104 accidents), occur on a road with two ways and a median island. This is followed by 35.5% of accidents occurring on a two-way road without a median island. The categories and their respective counts are presented in the dataset description file for the Road Type, Road Surface Description, and Road Properties attributes.

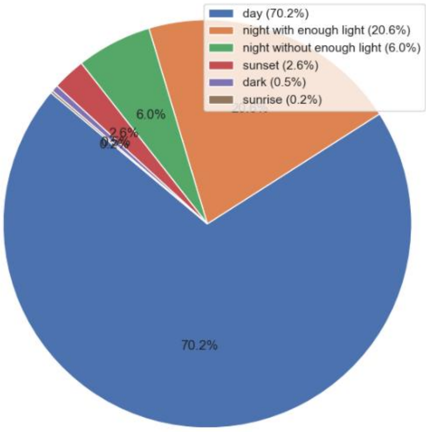

Figure 10. Distribution of Accidents by Daylight Types

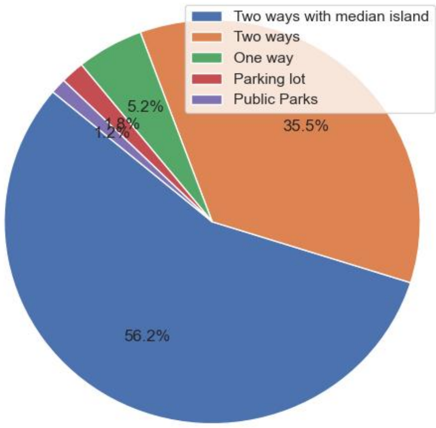

Figure 21. Distribution of Accidents by Road Lanes

3.1.5. Vehicle Features

The last features to consider are those related to the vehicles involved in accidents, such as vehicle type, the number of cars participating in the accidents, and vehicle nationality. The categorization of vehicle types is derived from the classification provided by the Directorate of Public Security-Driver and Vehicle Licensing Department in Jordan. Notably, passenger vehicles (up to 9 seats) account for most accidents, totaling 56,106 incidents.

3.2. Data Pre-processing

Text data was converted to numbers for data cleaning so that the models could process it more readily. The dataset did not contain any null values; hence, no null value processing was required. Another phase in the cleaning process was eliminating outliers. For example, values under 12 years old were eliminated from the driver age feature. Other extreme values were also eliminated.

The following step involves identifying the target feature; the dataset was evaluated to learn more about the data included and how it relates to earlier research. We decided to focus this study on accident severity, taking into account the quantity and kinds of injuries sustained as well as the number of fatalities per incident. We categorized accident severity into five classes: none, minor, moderate, severe, and critical. Figure 12 shows the distribution of accident severity across these five categories and the corresponding count for each category, highlighting a noticeable imbalance in the data.

Figure 13 illustrates the correlation between JO-TAD features and the target value, "Accident Severity." The correlation values are sorted by absolute value while preserving their signs. This illustration is important for identifying the most important features for predicting the target value.

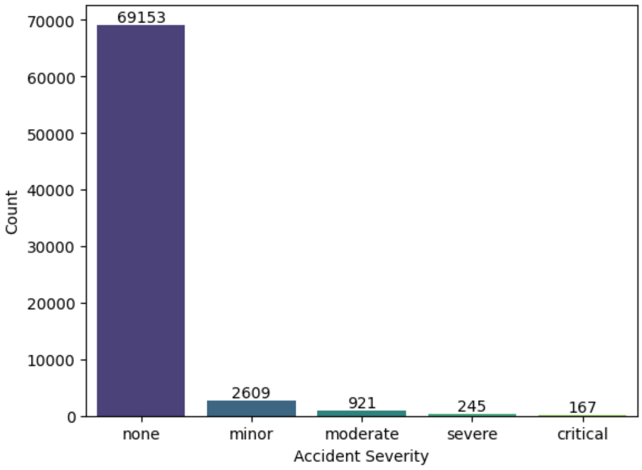

Figure 12. Distribution of accident severity across the five categories

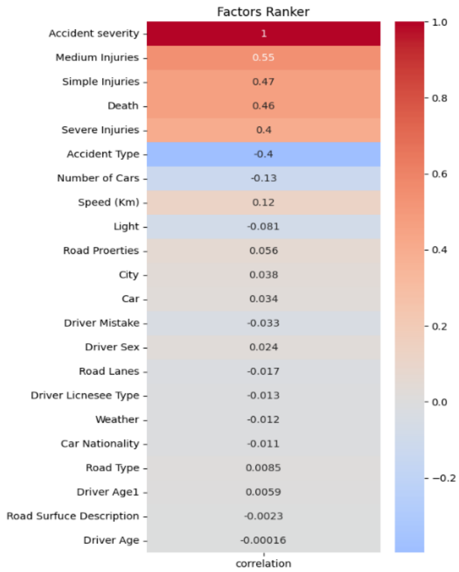

Figure 13. The features of the JO-TAD along with their correlations to the target value "Accident Severity"

3.3. Algorithms Used and Implementation

We focused on evaluating the performance of seven machine-learning models for accident severity prediction using commonly recommended best-practice parameters for each algorithm. While we did not perform explicit hyperparameter tuning, we ensured that the parameter values were selected based on established practices and widely accepted guidelines in the literature to ensure robust performance. The following subsections briefly describe ML algorithms used in this study [33].

3.3.1. Support Vector Machines (SVM)

SVM is a well-liked supervised learning technique for regression and classification applications. It can be used for both linear and non-linear classification with decision boundaries (hyperplane) and is efficient when working with high-dimensional data [34]. SVM operates by determining the most optimal hyperplane with the biggest margin that divides the data points of several classes. The margin is the distance, or support vectors, between the nearest data points from each class and the hyperplane. SVMs seek to increase this margin, which improves the model's resilience and generalization.

3.3.2. K-Nearest Neighbors (KNN)

KNN is a supervised learning method that is easy to use and efficient for regression and classification problems. The KNN method operates based on data point similarity and is non-parametric. It produces predictions based on the degree of similarity between the input data points and their nearest neighbors in the training set. Upon receiving a new data point, KNN searches the training dataset for the K points that are closest to it and then uses the weighted average of those neighbors' labels or the majority vote to predict the label of the new point [34].

3.3.3. Naive Bayes (NB)

NB is a straightforward yet powerful classification technique that relies on feature independence and the Bayes theorem. NB is a popular approach with a wide range of applications, despite its basic assumptions, making it very effective. Because it assumes that every feature is independent of every other feature, which is frequently not the case in real-world circumstances, NB is referred to as "naive." For many classification problems, NB can still be computationally efficient and successful, even with this simplifying assumption. The probabilistic algorithm NB bases its prediction process on the Bayes theorem [35].

3.3.4. Logistic regression (LR)

In machine learning and artificial intelligence (AI), LR is a kind of supervised learning algorithm utilized for binary classification tasks. It is derived from a collection of input variables and denotes one of two possible results, usually denoted by the numbers 0 and 1. A modification of the LR model, the multinomial logistic regression method predicts the probability distribution as a multinomial probability distribution and changes the loss function to a cross-entropy loss so that multi-class classification problems can be supported properly. Numerous industries, including banking, healthcare, and the social sciences, have extensively used LR.

3.3.5. eXtreme Gradient Boosting (XGBoost)

XGBoost is an extreme gradient boosting algorithm and decision tree-based ensemble learning model. It focuses on fitting and learning residuals from previous trees, resulting in a final prediction result [36], [37]. XGBoost is used by data scientists and researchers to optimize machine-learning models. It is efficient, scalable, and fast, with performance comparable to other state-of-the-art algorithms [38].

3.3.6. Random Forest (RF)

A machine learning system called Random Forest builds an ensemble of several decision trees to produce predictions that are more accurate. It is a collection of test sets consisting of classification and regression trees (CART) trained on bootstraps [39]. The method adheres to certain guidelines for post-processing, self-testing, tree combination, and tree growth [40], [41]. It is stable in high-dimensional parameter spaces and resistant to overfitting. Important hyperparameters are min_sample_leaf, max_features, and n_estimators. RF has many uses, works well for both regression and classification, and makes it simple to see how much weight is given to different input features. In terms of performance, it is difficult to surpass and a fantastic option for early model development. It is not without restrictions, though. Generally speaking, additional trees improve performance and stabilize predictions, but they also slow down computation.

3.3.7. Decision Tree (DT)

A non-parametric supervised learning technique for regression and classification is called a decision tree. After dividing the training data according to an attribute value restriction, the DT model applies a hierarchical decomposition to the training data space. The restriction or predicate depends on whether the word is present or absent. Until the leaf nodes have enough records to meet the classification objective, recursive data space splitting is done [34], [42].

4. Results and Discussion

This section explains in depth the experiments on the prediction models of accident severity and the findings from each experiment. It also covers the discussion of the results.

4.1. Evaluation Metrics

We employed precision, recall, and f1-score, which are computed for each of the five classes— critical, severe, moderate, minor and none—to assess the categorization prediction outcomes of each experiment. The outcomes are then averaged. Additionally, the overall classifier accuracy is assessed. An ML metric called accuracy calculates the percentage of accurate predictions a model makes relative to all of its predictions. It is among the most popular measures for assessing a classification model's performance. It offers a broad performance metric and is simple to compute and comprehend. Precision and recall are frequently more instructive than raw accuracy in our situation, when class imbalance is an issue, so we include them as well. Precision, which gauges the accuracy of positive predictions, is helpful when the cost of false positives is large. Recall, which gauges the capacity to spot positive cases, is crucial when the cost of false negatives is large. f1-score beneficial in datasets that are unbalanced; strikes a balance between recall and precision.

4.2. Results

A training and testing set was created using the data outlined in Section 3.1, with 75% of the data used for training and 25% for testing. We applied k-fold validation with k=5 on the whole train split, k-fold validation lessens the unpredictability brought on by any one train-test split by providing an average performance score (accuracy, F1-score, etc.) across all folds. As a result, the model performance estimate becomes more consistent and trustworthy. The seven classifiers under study were trained and tested through experiments. Table 3 provides an overview of all models' evaluation outcomes.

Table 3. Overview of all classifier models' Performance outcomes

Algorithm	Precision	Recall	F1-Score	Accuracy
SVM	92.75%	95.28 %	93.84%	95.28%
KNN	92.46%	94.31%	93.01%	94.31%
NB	89.42%	94.56%	91.92%	94.56%
LR	97.80%	98.10%	97.37%	98.10%
DT	92.74%	93.67%	93.16%	93.67%
RF	93.28%	95.02%	93.78%	95.02%
XGBoost	92.75%	95.27%	93.84%	95.27%

With 98.1% accuracy, 97.8% precision, 98.10% recall, and 97.37% F1-score, the multinomial LR performed the best out of all the implemented algorithms. It was a good performer across the board. Although DT's accuracy and recall scores are very low (93.67%), they can be raised in further work by tuning its hyperparameters. Both the precision (89.42%) and F1-score (91.92%) are lowest for NB.

Based on the data, the NB algorithm had the lowest precision of 89.42%, which suggests a comparatively low number of correctly identified instances. The best method for accident classification should be chosen after considering these evaluation criteria as well as additional elements like project needs, interpretability, and computing efficiency.

Figure 14, figure 15, figure 16, figure 17, figure 18, figure 19, and figure 20 show the confusion matrices of the seven classifiers (SVM, KNN, NB, LR, DT, XGBoost, and RF). The quantity and type of injuries are directly correlated with the five classes (0=none, 1=minor, 2=moderate, 3=severe, and 4=critical) of the JO-TAD. A significant number of records in the dataset had 0-class accidents, indicating an imbalance in the dataset. For example, the LR model performs exceptionally well for the "none" class, achieving (17,000+) true positives. This indicates high precision and recall for this class. The model struggles with minority classes ("minor," "severe," "critical"), leading to poor performance for these categories. This may be attributed to an imbalanced dataset or overlapping feature spaces among certain classes.

ML classifiers' relative performance can differ depending on the dataset's properties and the problem you are attempting to address. Different classifiers may perform better under different conditions and with different optimization techniques. KNN, for instance, works well in situations when the data show local patterns and the decision boundary is non-linear. SVM performs best when the data has a distinct margin of separation. It works well in situations with complicated decision boundaries and can handle high-dimensional spaces.

When working with high-dimensional feature spaces and text data, NB performs effectively. It is predicated on feature independence and is computationally efficient. Both category and numerical data can be handled using DT. It is comprehensible and may be used to solve situations with intricate decision boundaries. LR is a straightforward but effective strategy when there is a linear relationship between the target variable and the features. It can be easily recognized and put into effect for binary classification.

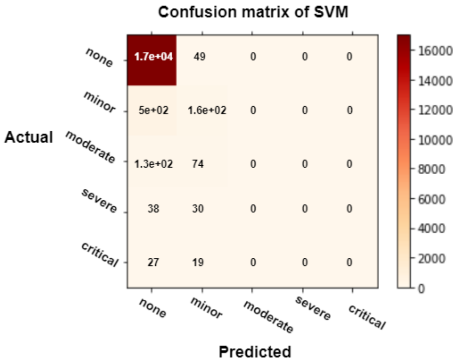

Figure 14. SVM algorithm: Confusion matrix

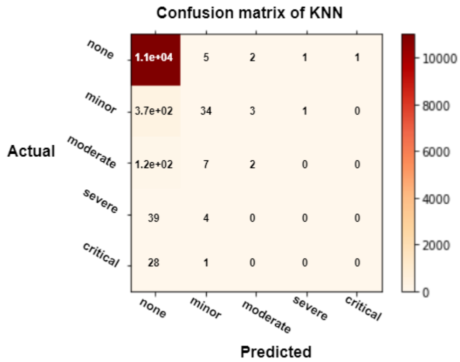

Figure 15. KNN algorithm: Confusion matrix

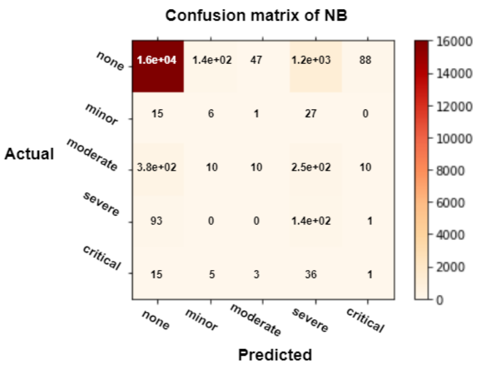

Figure 16. NB algorithm: Confusion matrix

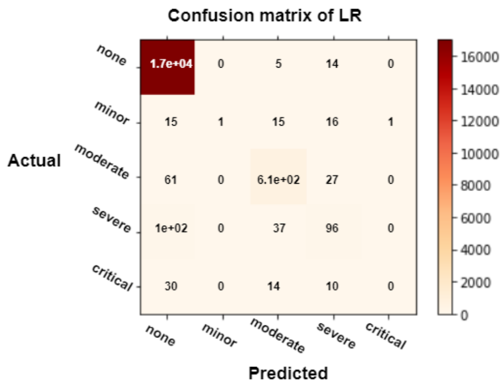

Figure 17. LR algorithm: Confusion matrix

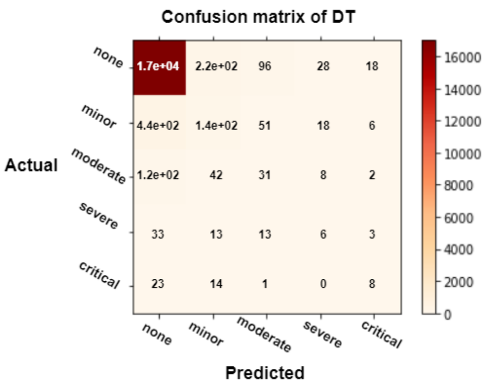

Figure 18. DT algorithm: Confusion matrix

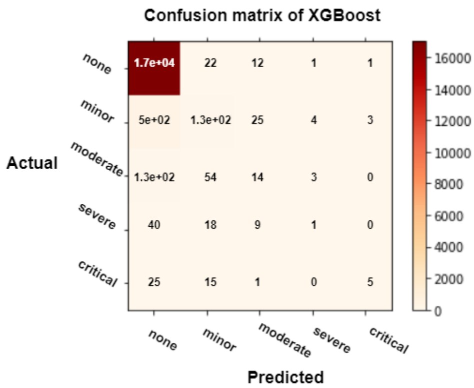

Figure 19. XGBoost algorithm: Confusion matrix

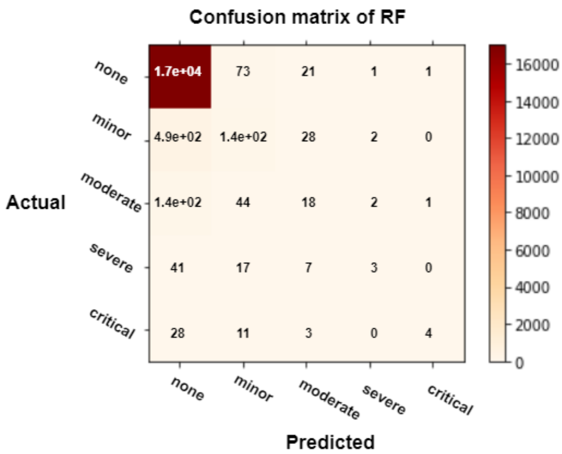

Figure 20. RF algorithm: Confusion matrix

The number of neighbors in KNN and the regularization parameter C in SVM are examples of hyperparameters that can be tuned. Other possible optimizations include feature scaling (SVM), ensemble methods like RF or gradient boosting to improve predictive performance, class balancing using oversampling or undersampling techniques, and overfitting prevention techniques like tree pruning for DT and L1 or L2 regularization for LR [43], [44].

Time efficiency is crucial, particularly for systems that foresee accidents. The results showed how computationally efficient each model was and gave useful information regarding whether or not Jordanian traffic management and accident severity prediction could be used. There were few differences in execution times for KNN, LR, SVM, NB, XGBoost, RF and DT. LR had the quickest execution times, whereas SVM and KNN took the longest.

4.3. Analysis of Results

This section focuses on explaining the significance of the model's findings, how the model's accuracy and other metrics translate to practical insights, and what these mean for stakeholders such as policymakers, traffic management authorities, and the general public.

The analysis of traffic accident severity in Jordan using machine learning models provides critical insights into the factors contributing to accident outcomes and the effectiveness of various predictive algorithms. The results indicate that LR emerged as the most accurate model with a 98.1% accuracy rate. This high accuracy suggests that LR is particularly effective in capturing the linear relationships between input features and accident severity. This could be due to effective feature engineering, L2 regularization, and straightforward relationships in the dataset. The preprocessing phase involved one-hot encoding and interaction terms, aligning with LR's assumption of linear relationships. L2 regularization improved robustness and generalization to test data, while more complex models like RF and XGBoost capture non-linear interactions.

The LR model's superior accuracy highlights its capability to handle categorical and continuous variables effectively, making it well-suited for datasets with diverse types of data inputs, such as those in accident records. This model's interpretability is a key advantage, allowing stakeholders to understand which factors most significantly influence accident severity.

Although LR was the most accurate, RF and KNN also performed well, with accuracies of 96.5% and 95.2%, respectively. These models are known for their robustness and ability to capture complex, non-linear interactions within the data. RF, in particular, provides valuable insights through feature importance, which can help identify the most critical variables influencing accident outcomes.

The models identified several significant predictors of accident severity, including vehicle speed, weather conditions, time of day, and road type. For instance, higher vehicle speeds were consistently associated with increased accident severity, which aligns with findings from previous research in other countries.

The emphasis on model interpretability is crucial for effective policy implementation. LR offers clear insights into the impact of each variable on accident severity, allowing policymakers to develop targeted interventions. For example, if certain road types are identified as high-risk, measures such as improved signage or speed limit enforcement can be prioritized.

In order to put safety improvement measures into place and avoid congestion, transportation experts and officials in Jordan should place a high priority on knowing when and where accidents occur on road networks. Techniques for locating dangerous roads should be based on a thorough and logically grounded knowledge of accident statistics, especially in countries with limited budgets. For instance, they can implement measures to lower traffic speed, improve road sign visibility, and maintain safe distances when they know where the majority of accidents (collisions) occur. Additionally, when the responsible authority is aware of the most common driving errors that result in accidents, they can take steps to stop them or use penalties and fees to make drivers observe traffic laws, including always wearing seat belts and other safety gear.

The practical application of these models can lead to more informed decision-making processes for traffic authorities. By integrating these predictive models into traffic management systems, authorities can proactively address potential risks and allocate resources more effectively to areas with higher predicted accident severity.

5. Conclusion

This study investigated the efficiency of different machine-learning algorithms in predicting the accidents fatality rates, obtained knowledge about current road accident status in Jordan, and learned more about the variables leading to their severity. Real traffic accident data from Jordan has been gathered, examined, and made accessible for future studies. This work has effectively assessed and compared the performance of various multiclass ML classifiers, such as the SVM, NB, KNN, LR, DT, XGBoost and RF algorithms. The most accurate algorithm was found to be Multinomial LR, which achieved an astounding accuracy of 98.1%. These results provide insightful information about the efficiency of ML algorithms in modelling accident severity prediction and can direct ongoing studies and applications in relevant fields. Future research will explore advanced techniques such as ensemble methods, deep learning, and feature selection to further enhance model performance. We will prioritize features with high correlation to accident severity, such as weather conditions, time of day, and road type, while addressing the challenge of class imbalance, particularly for critical and severe accidents.

The interpretability of the machine learning models used in this study is pivotal for translating data-driven insights into actionable strategies. By understanding the relationships between various factors and accident severity, stakeholders can develop more effective road safety measures, ultimately reducing the incidence and severity of traffic accidents in Jordan. This research underscores the value of transparent and interpretable machine learning models in public safety and policy-making contexts.

6. Declarations

6.1. Author Contributions

Conceptualization: D.M., and M.F.A.; Methodology: D.M. and S.M.K.; Software: D.M., S.M.K., and M.F.A.; Validation: D.M. and M.F.A.; Formal Analysis: D.M. and M.F.A.; Investigation: D.M.; Resources: D.M.; Data Curation: D.M.; Writing Original Draft Preparation: D.M., S.M.K., and M.F.A.; Writing Review and Editing: D.M., M.F.A., and S.M.K.; Visualization: M.F.A and S.M.K. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available at the Mendeley website (Link: https://doi.org/10.17632/r6db558376.1).

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A. Sartori, A. Russo, A. Sardo, and D. Raniero, "Fatal road traffic accidents and injuries: a preliminary study," *Int. J. Legal Med.*, vol. 139, no. 1, pp. 375–381, 2025.

[2] J. Schoner, R. Sanders, and T. Goddard, "Effects of advanced driver assistance systems on impact velocity and injury severity: an exploration of data from the crash investigation sampling system," *Transp. Res. Rec.*, vol. 2678, no. 5, pp. 451–462, 2024.

[3] H. Al-Masaeid, "Traffic accidents in Jordan." *Jordan Journal of Civil Engineering* vol. 3, no. 4, pp. 331-343, 2009.

[4] M. Labib, A. Rifat, M. Hossain, A. Das, and F. Nawrine, "Road Accident Analysis and Prediction of Accident Severity by Using Machine Learning in Bangladesh," *2019 7th International Conference on Smart Computing and Communications (ICSCC)*, Sarawak, Malaysia, vol. 2019, no. 6, pp. 1-5, 2019.

[5] X. Xu, B. Chen, and F. Gan, "Traffic safety evaluations based on grey systems theory and neural networks," *in Proc. WRI World Congress on Computer Science and Information Engineering*, vol. 5, no. 3, Los Angeles, CA, USA, pp. 603–607, 2009.

[6] A. Alqatawna, A. M. R. Álvarez, and S. S.-C. García-Moreno, "Comparison of multivariate regression models and artificial neural networks for prediction highway traffic accidents in Spain: a case study," *Transp. Res. Procedia*, vol. 58, no.1, pp. 277–284, 2021.

[7] D. Mustafa, L. Sawaged and S. M. Khabour,, "Predictive Analysis of Jordan Road Accidents Severity Using Machine Learning," in *2023 International Symposium on Networks, Computers and Communications (ISNCC)*, vol. 2023, no. 10, pp. 1–6, 2023.

[8] H. H. Pour, F. Li, L. Wegmeth, C. Trense, R. Doniec, M. Grzegorzek, and R. Wismüller, "A machine learning framework for automated accident detection based on multimodal sensors in cars," *Sensors*, vol. 22, no. 10, pp. 3634-3654, 2022.

[9] S. P. Ardakani, X. Liang, K. T. Mengistu, R. S. So, X. Wei, B. He, and A. Cheshmehzangi, "Road Car Accident Prediction Using a Machine-Learning-Enabled Data Analysis," *Sustainability*, vol. 15, no. 7, pp. 5939-5953, 2023.

[10] M. Balfaqih, S. A. Alharbi, M. Alzain, F. Alqurashi, and S. Almilad, "An accident detection and classification system using internet of things and machine learning towards smart city," Sustainability, vol. 14, no. 1, pp. 210-222, 2021.

[11] *P. Deitel and H. Deitel*, *Intro to Python® for Computer Science and Data Science: Learning to Program with AI*, Big Data and the Cloud, 1st ed. Pearson Education, 2022.

[12] T. Vaiyapuri and M. Gupta, "Traffic accident severity prediction and cognitive analysis using deep learning," Soft Computing, vol. 28, no. 2, pp. 1–13, 2021.

[13] N. Tamascelli, R. Solini, N. Paltrinieri, and V. Cozzani, "Learning from major accidents: A machine learning approach," *Computers and Chemical Engineering*, vol. 162, no. 6, pp.107786, 2022.

[14] A. Jamal, M. Zahid, M. T. Rahman, H. M. Al-Ahmadi, M. Almoshaogeh, D. Farooq, and M. Ahmad, "Injury severity prediction of traffic crashes with ensemble machine learning techniques: A comparative study," International Journal of Injury Control and Safety Promotion, vol. 28, no. 4, pp. 408–427, 2021.

[15] M. M. Chong, A. Abraham, and M. Paprzycki, "Traffic Accident Analysis Using Machine Learning Paradigms", Informatica (Slovenia), vol. 29, no. 1, pp. 89–98, May 2005.

[16] A. Atwah and A. Al-Mousa, "Car Accident Severity Classification Using Machine Learning," 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), Zallaq, Bahrain, vol. 2021, no. 9, pp. 186–192, 2021.

[17] A. Patel, A. Uniyal, and R. Dhanare, "U.K. Car Accident and Insurance Predictor using Machine Learning," in 2021 2nd Global Conference for Advancement in Technology (GCAT), Bangalore, India, vol. 2021, no. 10, pp. 1–6, 2021.

[18] S. S. Yassin and Pooja, "Road accident prediction and model interpretation using a hybrid K-means and random forest algorithm approach," *SN Applied Sciences*, vol. 2, no. 9, pp. 1-13, 2020.

[19] T. Bokaba, W. Doorsamy, and B. S. Paul, "Comparative Study of Machine Learning Classifiers for Modelling Road Traffic Accidents," *Applied Sciences*, vol. 12, no. 2, pp. 828-845, 2022.

[20] S. H. Hashmienejad and S. M. H. Hasheminejad, "Traffic accident severity prediction using a novel multi-objective genetic algorithm," *International Journal of Crashworthiness*, vol. 22, no. 4, pp. 425–440, 2017.

[21] M. Yan and Y. Shen, "Traffic Accident Severity Prediction Based on Random Forest," *Sustainability*, vol. 14, no. 3, pp. 1729-1741, 2022.

[22] J. H. Yousif and M. S. AlRababaa, "Neural technique for predicting traffic accidents in Jordan," *Journal of American Science*, vol. 9, no. 11, pp. 528–525, 2013.

[23] M. I. Sameen and B. Pradhan, "Severity prediction of traffic accidents with recurrent neural networks," *Applied Sciences*, vol. 7, no. 6, pp. 476-492, 2017.

[24] B. Geyik and M. Kara, "Severity prediction with machine learning methods," *2020 International Congress on Human-Computer Interaction, Optimization, and Robotic Applications (HORA)*, Ankara, Turkey, vol. 2020, no. 6, pp. 1–7, 2020.

[25] N. AlKofahi, "The Effects of Refugee at Traffic Accidents in Jordan," *International Journal of Engineering and Technology (IJET)*, vol. 12, no. 4, pp. 517–526, Aug. 2020.

[26] X. Wen, Y. Xie, L. Jiang, Z. Pu, and T. Ge, "Applications of machine learning methods in traffic crash severity modelling: Current status and future directions," *Transp. Rev.*, vol. 41, no. 6, pp. 855–879, 2021,

[27] M. E. Shaik, M. M. Islam, and Q. S. Hossain, "A review on neural network techniques for the prediction of road traffic accident severity," *Asian Transport Studies*, vol. 7, no. 2021, pp. 1-11, 2021.

[28] S. Tayebi Khorami and E. Sacchi, "Validation of machine learning algorithms as predictive tools in the road safety management process: Case of network screening," *Journal of Transportation Engineering, Part A: Systems*, vol. 148, no. 9, pp. 1-21, 2022.

[29] Z. Sun, D. Wang, X. Gu, M. Abdel-Aty, Y. Xing, J. Wang, ... and Y. Chen, "A hybrid approach of random forest and random parameters logit model of injury severity modeling of vulnerable road users involved crashes," *Accident Analysis and Prevention*, vol. 192, no. 11, pp.1-18, 2023.

[30] Y. Yang, K. Wang, Z. Yuan, and D. Liu, "Predicting freeway traffic crash severity using XGBoost-Bayesian network model with consideration of features interaction," *J. Adv. Transp*., vol. 2022, no. 4, pp. 1-15, 2022

[31] H. Zhao, H. Yu, D. Li, T. Mao and H. Zhu, "Vehicle Accident Risk Prediction Based on AdaBoost-SO in VANETs," in *IEEE Access*, vol. 7, no. 1, pp. 14549-14557, 2019.

[32] D. R. Mayhew, S. A. Ferguson, K. J. Desmond, and G. M. Simpson, "Trends in fatal crashes involving female drivers, 1975–1998," *Accident Analysis and Prevention*, vol. 35, no. 3, pp. 407–415, 2003.

[33] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011.

[34] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*. Cambridge, UK: Cambridge Univ. Press, vol. 2014, no. 5, pp. 1-12, 2014.

[35] S. Malik, H. El Sayed, M. A. Khan, and M. J. Khan, "Road accident severity prediction—a comparative analysis of machine learning algorithms," *in 2021 IEEE Global Conference on Artificial Intelligence and Internet of Things (GCAIoT)*, Dubai, UAE, vol. 2021, no. 12, pp. 69–74, 2021.

[36] P. Lu, Z. Zheng, Y. Ren, X. Zhou, A. Keramati, D. Tolliver, and Y. Huang, "A gradient boosting crash prediction approach for highway-rail grade crossing crash analysis," *Journal of Advanced Transportation*, vol. 2020, no. 1, pp. 1-10, 2020.

[37] I. Altaf and A. Kaul, "Classifying collisions in road accidents using XGBoost, CatBoost and SALP Swarm-based optimization algorithms," *Multimedia Tools and Applications*, vol. 83, no. 13, pp. 38387–38410, 2024.

[38] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD '16)*, vol. 2016, no. 8, New York, NY, USA, pp. 785–794, 2016.

[39] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, Classification and Regression Trees. Boca Raton, FL, USA: CRC Press, vol, 2017, no. 10, pp. 1-12, 2017.

[40] R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms," in 23rd Int. Conf. *Machine Learning (ICML), Pittsburgh, PA*, vol. 2006, no. 6, pp. 161–168, 2006.

[41] B. H. Menze, B. M. Kelm, R. Masuch, U. Himmelreich, P. Bachert, W. Petrich, and F. A. Hamprecht, "A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data," *BMC Bioinformatics*, vol. 10, no. 12, pp. 1-16, 2009.

[42] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093–1113, 2014.

[43] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, vol. 415, no, 11, pp. 295–316, 2020.

[44] S. Sun, Z. Cao, H. Zhu, and J. Zhao, "A survey of optimization methods from a machine learning perspective," *IEEE Transactions on Cybernetics*, vol. 50, no. 8, pp. 3668–3681, 2020.