Available online at www.CivileJournal.org

Civil Engineering Journal

(E-ISSN: 2476-3055; ISSN: 2676-6957)

Vol. 10, No. 06, June, 2024

Utilizing GIS and Machine Learning for Traffic Accident Prediction in Urban Environment

Atif Ali Khan [1*], Jawad Hussain [1]

[1] *Department of Civil Engineering, University of Engineering and Technology Taxila, HMC Link Road Taxila, 47050, Pakistan.*

Received 22 January 2024; Revised 19 May 2024; Accepted 25 May 2024; Published 01 June 2024

Abstract

Traffic accident prediction is crucial to preventive measures against accidents and effective traffic management. Identifying hotspots can facilitate the selection of the most critical survey points to note the contributing features. In this research, an effort has been made to identify hotspots and predict traffic accident occurrences in an urban area. Accident data was obtained from the Rescue 1122 Emergency Services of Faisalabad, and hotspots were identified using Moran's I in ArcGIS. Results showed that most hotspots were located around the General Transport Stand (GTS) area due to the maximum number of road users. The temporal investigations showed that the accident occurrence was significant from 1 to 2 p.m. The identified hotspots were further investigated by conducting a field survey. Essential features such as road geometric features, road furniture, and traffic data were used for developing Machine Learning Algorithms for accident prediction. Using Computer Vision, traffic data was extracted from recorded videos. Random forest, linear regression, and Decision tree algorithms were developed using Python in the Jupyter Notebook environment. The decision tree algorithm showed a maximum accuracy of 84.4%. The analysis of contributing factors revealed that road measurements had the maximum effect on accident occurrence.

Keywords: ArcGIS; Hotspots Analysis; Local Moran's I Static; Spatial Analysis; Road Traffic Accidents; Machine Learning; Traffic Accident Prediction.

1. Introduction

Road Traffic Accidents are one of the leading causes of death all around the world. Moreover, road accidents result in injuries at different severity levels if not found fatal. Both of these types of accidents cause financial losses on a larger scale. The financial damage from road accidents harms the economies of developing and under-developed countries at higher grades. This might be associated with these countries being densely populated, and the quality of the available infrastructure facilities needs to be on par with the developed countries. According to the WHO, approximately 1.3 million individuals die in traffic accidents worldwide each year. 93% of these fatalities occur in low- and middle-income countries, despite these nations having approximately 60% of the world's vehicles. Traffic accidents cost approximately 3% of the GDP [1]. Pakistan is one of the developing countries currently trying to develop and improve its infrastructure for the betterment of its citizens. These developments are primarily concentrated in the major cities. This results in the rapid growth of cities and their urban areas as people migrate towards the cities to gain access to better facilities. Pakistan, being one of the most populist countries in the world, has a population of 241.49 million, according to the 2023 census. Among the four provinces of Pakistan, the population of Punjab is 127.68 million, the highest among all the provinces. In 2017, about 36.86% of the Punjab population lived in urban areas, which increased to 40.70% according

* Corresponding author: atif.ali@iefr.edu.pk

http://dx.doi.org/10.28991/CEJ-2024-010-06-013

to the census of 2023. Faisalabad district is the third largest district by population in Pakistan and the second largest in the eastern province of Punjab, with a population of 9,075,819 in 2023 [2]. Due to urbanization, various areas of the city are becoming densely populated, resulting in an increase in traffic and accidents. According to the World Health Organization (WHO), the death rate in Pakistan resulting from road traffic accidents is 15.18 per 100,000 people, which is very high [3, 4]. Therefore, there is a need for interventive measures to reduce this rate. It is important to identify the areas with a high frequency of accidental occurrences along with the corresponding contributing factors for reducing accident occurrences efficiently and effectively. Areas with a high frequency of accidents are called hotspots. The most economical and effective way to identify hotspots is through the use of Geographic Information Systems (GIS). Various GIS programs have been developed to identify hotspots, such as ArcGIS, QGIS, etc. [5–10]. These programs utilize different techniques to analyze the statistical significance of hotspot locations, such as Local Moran's I, KDE, Getis Orid G*, etc., which are widely used. Furthermore, machine learning techniques are becoming popular in the prediction of accidents. The combination of GIS and machine learning will allow the researchers to not only identify the hotspots but to evaluate associated contributing factors as well. This will help in the identification of different factors that affect the occurrence of accidents in the hotspots. This information can then be used effectively for targeted implementation in hotspots. In this research, an effort has been made to develop a methodology combining hotspot identification using GIS technology and accident prediction using machine learning.

Geographical information systems (GIS) have become a valuable tool to identify and analyze traffic accidents occurring on a road network. A Geographic Information System (GIS) is a comprehensive system that generates, oversees, and assesses various forms of data. It merges location information (spatial data) and diverse descriptive data (attributes) with maps [11]. Different estimation and zoning techniques help in the identification of hotspots by clustering the data points on the map. These hotspots can be analyzed based on their statistical significance level. The highly significant hotspots can be prioritized for the application of remedial measures. For this purpose, these identified hotspots are further evaluated to explore causes that contribute to accident occurrence. This methodology helps in the evaluation of a road network in all types of environments, such as urban, rural, metropolitan, suburban, etc. In a study, Khatun et al. (2024) used GIS techniques to identify hotspots using Kernel Density Estimation on national highways in Bangladesh [6]. The Kernel density estimation technique has also been used by various researchers for the identification of traffic accident hotspots [12–15]. In an urban environment, there are different locations where the probability of an accident occurring is very high. Examples of these locations can be schools, offices, public buildings, etc. One such study was conducted by Chun et al. (2024), in which the hotspot formulation inside and outside of the school zones using Getis-Ord Gi* and Local Moran's I techniques was investigated [7].

Apart from the identification of hotspots, various other analyses are conducted to explore various trends in the data. These analyses are helpful in the classification of accidents according to various types, such as accident occurrence, accident severity, and injury. Spatial accident distribution can follow a specific pattern or be distributed randomly. Afolayan et al. (2022) identified accident hotspots along Nigeria's Lokoja-Abuja-Kaduna highway (2013–2017) using GIS techniques. In this research, it was found that the spatial autocorrelation analysis revealed a generally random accident distribution [16]. In another study, Krickovic et al. (2024) applied the Getis-Ord Gi* and Mann Kendall techniques to the ArcGIS Pro program to identify hotspots. Trend analysis was also carried out to investigate trends in the number of traffic accidents, traffic accident fatalities, and injuries [17]. The identified locations and factors need to be addressed to increase road safety at those locations. For example, in some cases, different amendments in road design and land use can help in a significant reduction in accident occurrence. Umair et al. (2022) studied urban design factors influencing road traffic crashes in developing countries. They applied Kernel Density Estimation technique on crash data and identified traffic accident hotspots. Key factors such as land use, street furniture, and road conditions were found to affect accident density [18].

Various GIS techniques allow researchers to cluster the zones based on the occurrence and severity of accidents. It is very important to classify road accidents in terms of severity, as zones with higher severity levels will require higher priority than zones with lower severity levels. Statistical techniques associated with GIS help in computing the statistical significance of zones with higher severity levels. The Getis Ord Gi* technique has been used by multiple researchers in the past to identify zones with higher severity levels of road accidents [15, 19–24]. In another study, Srikanth & Srikanth (2020) used Kernel Density Estimation (KDE) and Hotspot Analysis (Getis-Ord Gi*) to identify and rank accident hotspots in Des Moines City, United States, using crash data from 2008 to 2012. KDE visualizes hotspots, while Getis-Ord Gi* assesses statistical significance [20].

Local Moran's I is another statistical technique that is very popular among researchers for the identification of hotspots. The effects of different factors, such as severity, type of collision, type of road users, etc., were also investigated [21, 25, 26]. In one such study, Qalb et al. (2023) developed a semi-automated system using GIS and used statistical tools (Getis-Ord Gi* statistic, Moran's I spatial auto-correlation) to analyze and predict road crash patterns in Kasur, Pakistan. The study highlighted peak accident months, hours, and hotspots in the city, offering valuable insights for emergency response and road safety improvements [27]. Various seasonal and temporal factors have also been found to have an effect on accident occurrence. Erenler & Gümüş (2019) investigated Turkey road traffic accidents (RTAs) from 2013 to 2017, revealing that 697,957 accidents occurred, resulting in 1,168,121 injuries and 3,534 deaths. RTAs are more frequent on weekends, in the summer months, and in rural areas. Male individuals and motorcyclists face higher risks [28].

Different design attributes of road infrastructure, if not designed properly, can also contribute to accident occurrences. Certain behaviors are required of road users while traveling on certain types of roads. In urban areas, it is important to provide facilities for all types of road users so that they can complete their journeys safely. With the increase in urban population, existing road facilities are not able to facilitate the increasing number of road users. Failure of the road designs can result in the formulation of hotspots in specific zones. Statistical and GIS-based analyses identify hotspot zones using attributes like severity, road type, and lighting [29]. Hazaymeh et al. (2022) examined traffic accidents in Irbid Governorate, Jordan, from 2015 to 2019. Spatial autocorrelation and local hotspot analysis reveal clustered patterns of accidents along internal and arterial road networks, indicating the need for targeted safety measures. A 38% increase in reported accidents was observed, particularly in areas with high traffic volumes [30].

Various previous research has shown that the identification of hotspot areas plays an important role in the improvement of traffic safety by targeting remedial measures in the hotspot areas. Wang et al. (2022) utilized Optimized Hot Spot Analysis (OHSA) with varying optimal scales to identify traffic accident blackspots, considering different accident types and severity. Findings highlight central 'cities' higher safety challenges (54.6% accidents), guiding targeted interventions for improved road safety [31–33]. Rahman et al. (2020) used GIS analysis and geographically weighted regression (GWR) to uncover spatial relationships in Dammam, Saudi Arabia. In their study, they identified crash hotspots, crash causes (e.g., "sudden lane deviation"), and crash types ("collisions between motor vehicles") [10]. Hammas & Al-Modayan (2022) used geographic information systems to analyze traffic accidents in Medina, Menorah City. They identified accident hotspots, causes of the accidents, and safety measures to minimize traffic accidents [8].

Studies have shown that hotspot identification and spatial analysis can also help in investigating the effects of transportation infrastructure on accidents. Alkhadour et al. (2021) focused on road traffic safety in Amman, analyzing accidents temporally and spatially [29].

After the identification of traffic accident hotspots, these points can be further investigated to identify and mitigate the factors contributing to traffic accidents. Machine learning techniques have proven themselves to be efficient and effective methods to predict traffic accidents and the impact of associated contributing factors such as age, gender, driver behavior, road infrastructure/geometry, time, day, month, and year. Machine learning algorithms identify patterns and relationships in data, which help in the development of traffic accident prediction models.

Various past studies have used machine learning algorithms to analyze traffic data for accident prediction. Some of these techniques are decision tree (DT), support vector machine (SVM), k-nearest neighbor (k-NN), random forest (RF), logistic regression, Ridge Regression Naïve Bayes, and artificial neural networks (ANN). These machine algorithms provide different results in terms of accuracy when applied to the prediction of traffic accidents' occurrence, accident severity, and injury level. The variations in results occur due to differences in data type. Various contributing factors incorporated into the data provide insights into their impact on accident prediction. These factors are related to infrastructure (such as road type, road geometry, road intersection, number of lanes, etc.), environmental factors (such as weather conditions and light conditions), driver behavior, vehicle characteristics, and traffic data (such as average annual daily traffic, design speed, etc.).

For accident prediction, studies have shown that Random Forest models have provided the most accurate results among other machine learning algorithms. For example, Vyshnavi and Nalini (2022) used Random Forest, logistic regression, and decision tree algorithms for traffic accident prediction. The Random forest model provided the maximum accuracy and sensitivity out of the selected algorithms [34]. Similarly, Venkat et al. (2020) compared ANN, support vector machine, genetic algorithm, random forest, decision tree, KNN, and logistic regression for accident prediction. Attributes such as time, day, engine capacity, age of the vehicle, and weather conditions were used for developing the prediction models. It was concluded that Random forest model was the most accurate, with 96% [35]. In another study, Megibaru & Atnafu (2019) designed a machine-learning model to identify and predict accidents in Ethiopia using a random tree classifier. Attributes such as number of crashes, day of week, type of severity, road type, road geometry, road intersection, weather condition, light condition, cause of accident, and road condition were used in it. Random forest classifier showed maximum accuracy [36]. The study by Agustin Guerra et al. also found that the random forest model was the best to predict accident occurrences with the highest accuracy. Guerra et al. (2022) predicted crashes on road segments using an artificial neural network (ANN), random forest, support vector machine, linear regression, and generalized additive model (GAM). Graphical and equational comparisons were made in it. Attributes like route, zone, average annual daily traffic, design speed, and number of lanes were included [37]. Dogru & Subasi (2018) introduced an intelligent traffic accident detection system using vehicular ad-hoc networks (VANETs) to exchange vehicle data and issue alerts to drivers. The researchers employed machine learning (ML) techniques (Artificial Neural Networks, Support Vector Machines, and Random Forests), and the results showed Random Forests' superior performance, achieving 91.56% accuracy in accident detection [38]. Sripuram et al. (2022) compared random forest, decision tree, and logistic regression in traffic accident prediction. Random forest was found to be the best-suited model, and it gave accuracy above 80%. Vehicle reference, road class/type, location/longitude/latitude, light condition, weather condition, and junctions were the contributing factors to traffic accidents [39]. Ulu et al. (2024) used a geohash-based approach for traffic accident prediction. Decision tree (DT), k-nearest neighbor (k-NN), random forest (RF), and support vector machine (SVM) algorithms were applied to identify accident occurrence contributing factors in identified geohash regions. The results augment the use of machine learning models in forecasting accident occurrences using geohash-based models [40]. Sridevi et al. (2020) developed a model to predict accident-prone areas, considering various factors causing accidents. This model used data mining techniques (apriori) and machine learning concepts (K-means) to identify the factors causing accidents [41].

The selection and accuracy of a machine learning model depend on the data type and the selected features. Azimjonov et al. (2023) predicted accidents based on different contributing factors using machine learning algorithms. Various classifiers, including decision trees (J48, Random Forest, Rep Tree) and Bayesian classifiers (Naïve Bayes, Bayesian Network), were employed. Results indicated that J48 and Rep Tree classifiers had similar high F-measures (97.87% and 97.80%, respectively), while Random Forest performed slightly lower (90.9%) [42]. Torres & Asor (2021) employed data mining and classification algorithms to craft a predictive model for anticipating accident occurrences. Decision trees, k-nearest neighbor (K-NN), naïve Bayes, and neural networks were evaluated. Impressively, neural networks demonstrated high accuracy at 87.63%, excelling in accident classification. whereas K-NN and naïve Bayes achieved an accuracy of over 80% [43]. Ruzicka et al. (2019) predicted traffic accidents using neural networks. Vehicle type, weather condition, location, visibility, traffic intensity, and road condition were the attributes investigated in their research [44]. Pusuluri & Dangeti (2024) used QGIS for the identification of traffic accident blackspots in Visakhapatnam. Further, Machine Learning Algorithm DBSCAN was applied to identify crash clusters and validate the coordinates of the crash spots [5].

Machine Learning algorithms are being used to predict Traffic Accident Severity. This can help the authorities prioritize the most critical locations for timely interventions, depending on the predicted severity of accidents. Ulu et al. (2024) identified the primary contributing factors to traffic accident severity. The researchers employed a hybrid K-means and random forest approach. The random forest model predicted severity, achieving an impressive 99.86% accuracy, surpassing other classification techniques. Jadhav et al. (2020) evaluated the performance of the proposed approaches on the prediction of accident severity. The researchers determined the performance of each algorithm for four accident severity classes (Fatal / Grievous /Simple Injury/ Motor Collision). Naïve Bayes and Ada-Boost both achieve high accuracy among these four approaches, and their accuracy is 80%. While KNN and Decision Tree were found to be less accurate, overall, Ada-Boost gave the finest outcome due to its iterative classification on the Decision Tree [45]. Santos et al. (2021) applied supervised machine learning algorithms such as Decision Trees (DT), Random Forests (RF), logistic Regression (LR), and Naive Bayes (NB). They also applied unsupervised machine learning techniques, including DBSCAN and hierarchical clustering, for the prediction of accident severity. The findings showed that a rule-based model using the C5.0 algorithm (one of the algorithms used in the study) was more capable of accurately detecting the most relevant factors describing road accident severity [46].

Machine Learning algorithms can be helpful in the prediction of injury in a traffic accident. This can help authorities devise an effective rescue plan in an emergency. Candefjord et al. (2021) assessed machine learning algorithms like Logistic Regression, Ridge Regression, Bernoulli Naïve Bayes, Stochastic Gradient Descent, and Artificial Neural Networks in accident injury prediction. Logistic Regression emerged as the best performer, achieving an AUC of 0.86, indicating strong predictive ability. Factors like ejection, entrapment, belt use, airbag deployment, and crash type proved to be significant predictors [47]. Employing multiple logistic regression and artificial neural networks (ANN) as a machine learning approach, Najafi Moghaddam Gilani et al. (2021) identified influential variables affecting injury severity, fatal accidents, and property damage only (PDO). The multiple logistic regression method concluded with 89.17% prediction accuracy. Environmental factors like poor lighting, adverse weather, and vehicle quality also played significant roles in traffic accident severity. The ANN model outperformed, giving a 98.9% prediction accuracy, confirming its superior predictive capabilities for future accidents [48]. Driver experience, day and light conditions, driver age, and vehicle service year emerged as pivotal factors for different severity levels of traffic accidents [49]. In another study, Vipul Rana et al. (2019) predicted traffic accidents with the help of a logistic regression algorithm. The research concluded that road accidents are hugely affected by factors such as the types of vehicles, age of the driver, age of the vehicle, weather conditions, and road structure [50].

Pakistan is a developing country with limited budget allocation to traffic safety projects. Thus, targeted interventions in the identified accident-prone areas can prove economical. This can be achieved through hotspot identification using GIS technology. After identification of these areas, accident-contributing factors can be explored, and the effects of these factors on accident occurrence can be studied using Machine Learning. The results of Machine Learning can predict the significance of contributing factors, which can further provide recommendations for remedial measures.

In light of previous work carried out in a similar area, hotspots have been identified using Local Moran's I technique in ArcGIS. After this, Machine Learning algorithms have been used for further investigation of various factors contributing to the accident occurrence. These machine learning algorithms have been compared in terms of their accuracy in predicting accidents to identify the most suitable algorithm for the data type used in this study. The main objective of this research is accident prediction in an urban area using Machine Learning Algorithms. To achieve this, the following sub-objectives are formulated:

- Analysis of Traffic Accidents in an urban area;
- Identification of Traffic Accident hotspots in an urban area;
- Identification of significant factors contributing to Accidents.

2. Research Methodology

Pakistan is situated in southeast Asia which shares borders with Iran, Afghanistan, India, and China. The Arabian Sea is located in the south of Pakistan. Pakistan is divided into four provinces, with Punjab being its most populated province. Faisalabad is situated in the northeast of Pakistan, in the Punjab province. The urban area of Faisalabad was selected for this research and is shown in Figure 1. The boundaries enclose the urban region of Faisalabad city observed on the open street map at these coordinates: Point A (31.482137, 72.982303) near Aminpur Bypass Road, Point B (31.481898, 73.185003) near Faisalabad-Sheikhupura Road, Point C (31.343813, 72.982542) near College Road / Risalewala Road, and Point D (31.343574, 73.185721) near Satiyana Bypass Road in decimal degrees.

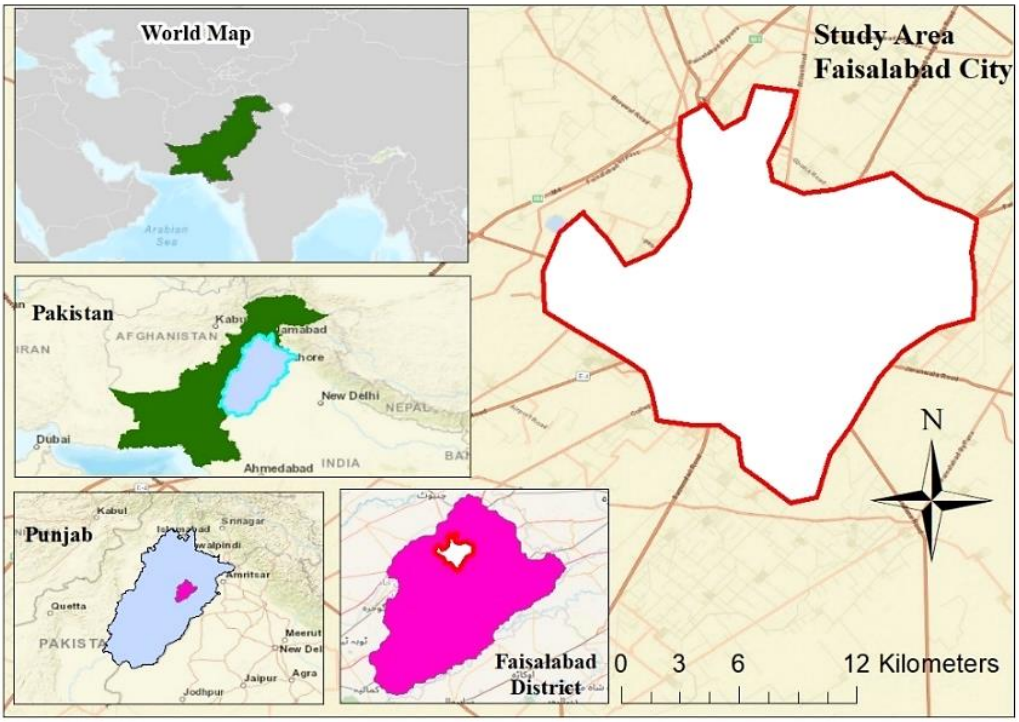

Figure 1. Study Area of Faisalabad

The methodology of the research is shown in Figure 2. The study area includes the urban area of Faisalabad, as explained above. The next step in this research was to collect Road Traffic Accident data, which was obtained from the Rescue 1122 Emergency Services of Faisalabad, Punjab, Pakistan. It contained attributes such as location, date, day, and time of the traffic crashes of the year 2022. Data screening was done by eliminating outliers such as points located outside of the study area, duplicate values, and missing entries. The screened data is the foundation for the spatial analysis conducted using ArcGIS. ArcGIS is a geographic information system (GIS) software suite developed by ESRI and has been used in this research. It allows users to collect, manage, analyze, and visualize spatial data, enabling them to make informed decisions based on location-based information [51–53]. The screened data was added to ArcGIS for performing spatial analysis. From previous work, the most suitable technique related to the obtained data type was found to be Moran's I static [54–56]. This method helps to identify accident hotspots by detecting spatial clustering and evaluating the significance of these hotspots. Local Moran's I Static begins by collecting data on traffic accidents, including their locations. The method then examines the spatial relationship between accident occurrences at different locations. It calculates a spatial autocorrelation coefficient called Moran's I [21].

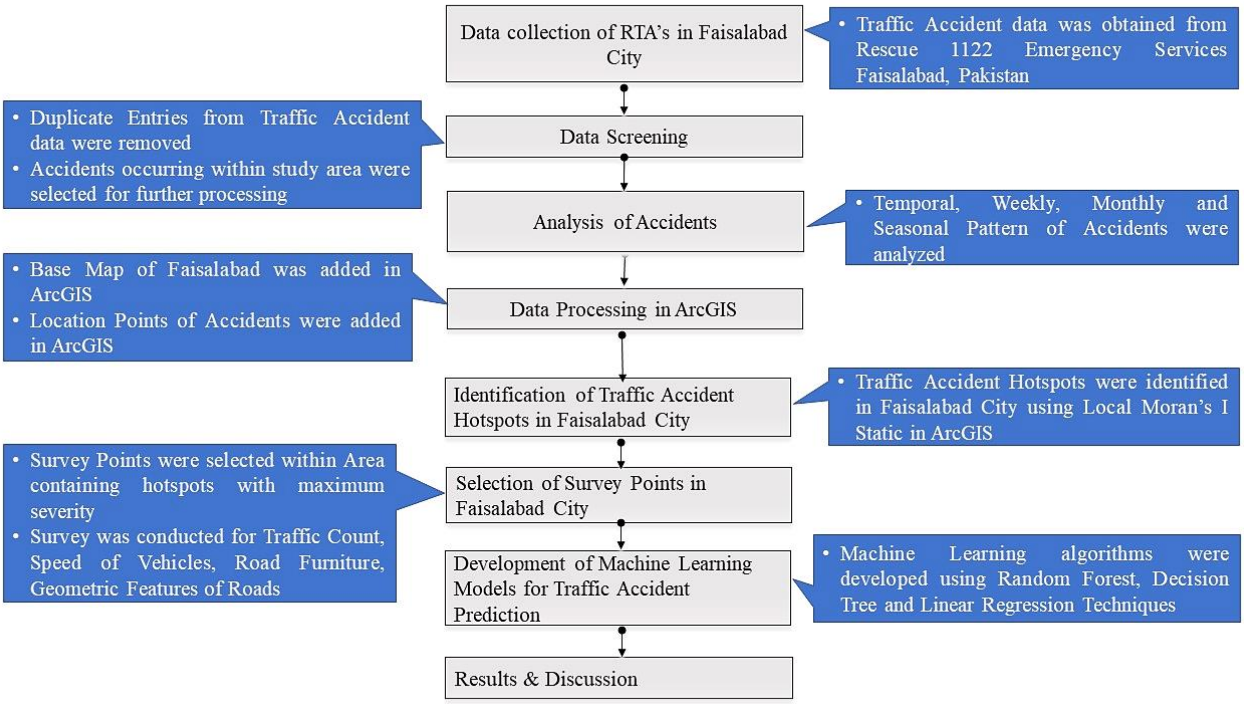

Figure 2. Research Methodology

After identification of the hotspot areas using Local Moran's I Static technique, a comprehensive effort was made to select survey points for obtaining traffic data and geometric features of roads and road furniture. From the results of ArcGIS, hotspots with higher significance levels were identified. Among these identified hotspots, seven survey points were carefully selected to represent all types of environments, traffic flow and composition, road geometry, and furniture. The environment type included the school, hospital, residential, and commercial areas of the study area. The traffic composition included all types of road users in the study area. Road geometry included features such as lane width, number of lanes, intersections, shoulders, median, etc. Road furniture included sign boards, traffic lights, etc. These selected points included A (31.4210, 73.0789) in Chiniot Bazaar, B (31.4181, 73.0971) at Railway Station, C (31.3991, 73.1523) on Jaranwala Road, D (31.4305, 73.1203) on Canal Road, E (31.4153, 73.0915) at General Transport Stand, F (31.4274, 73.1231) on Susan Road, and G (31.3955, 73.1143) on Satyana Road, as shown in Figure 3. These points have been shown on a map using Google Earth Pro.

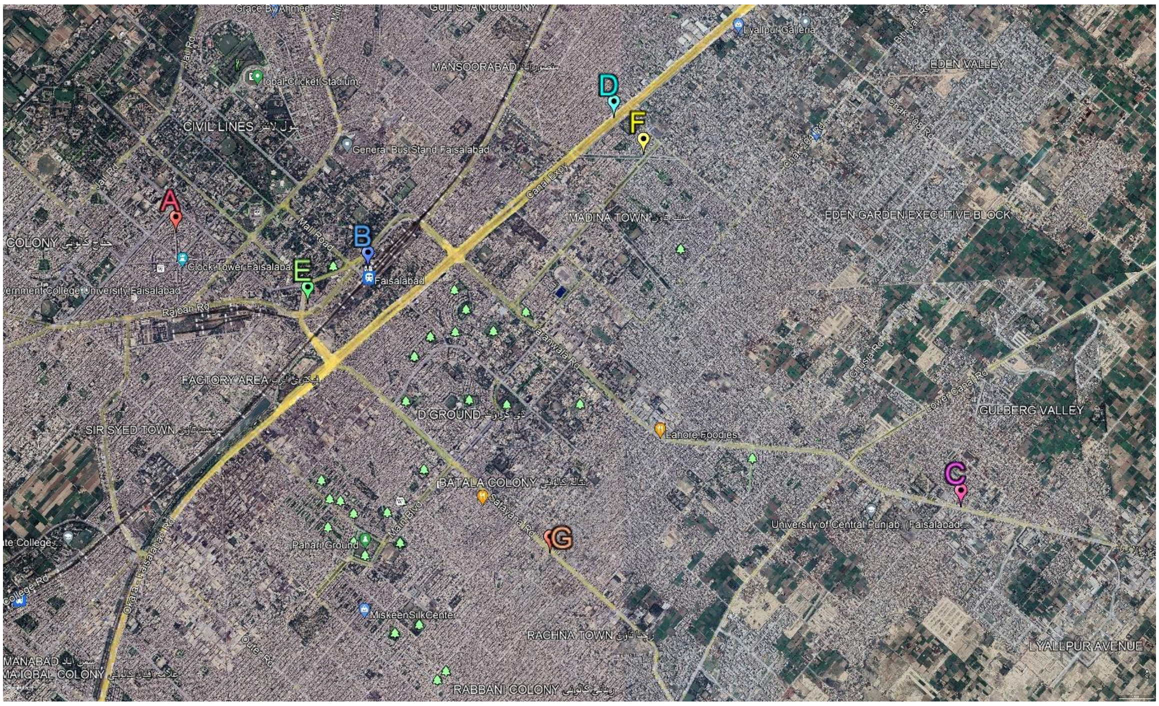

Figure 3. Selected Survey Points

The field survey was conducted at the above-mentioned points with the approval of the City Traffic Police Faisalabad. At these points, videos of the road traffic were recorded for traffic information like traffic flow, speed of vehicles, nature of traffic, maximum traffic flow at a specific time, No. of pedestrians crossing the roads. To extract the data on vehicle, count and speed from the recorded videos, the Python code of Computer Vision 2 (CV2) was used in Visual Studio Code. The program was trained in a way that counts cars, buses, trucks, bicycles, bikes, and other vehicles separately to know how many of them cross the road at selected time slots along with their speed. Geometric and infrastructure features of the roads at selected points were also noted. These include lane green belt width, diversion in the selected road section, Road measurements (Lane width, Road cross section width), Number of traffic lights, Number of turns, and public buildings (school, college). The data on traffic accidents, traffic speed, traffic vehicles, road infrastructure, and geometric features of roads was used for developing the Machine Learning Algorithm. The selected percentage of 80% (0.8) from the input data was designated for training the model. While the remaining 20% (0.2) was selected for the testing purpose [43].

The correlation of all noted attributes with the road traffic accident was developed using Python in the Jupyter Notebook environment. This correlation is shown in Table 4. Linear Regression, Random Forest, and Decision Tree machine learning algorithms were selected for this research as suggested by previous works for similar types of analysis. The codes for the aforementioned algorithms were prepared using the Python programming language in Jupyter Notebook. After this, the developed codes were evaluated with the help of the root mean square error in the accident prediction. The cross-validation of the data was done to estimate the performance of the built model on unseen data.

3. Results and Discussion

This research was conducted with the main objective of predicting traffic accidents in urban areas. For this, the most significant hotspots were identified using ArcGIS. Then survey points were selected for the contributing features used in developing the machine learning algorithms.

3.1. Analysis of Accidents

A descriptive analysis of the traffic accident data was carried out. Some specific parameters were observed and analyzed. One Way ANOVA test was performed in SPSS to compare the effect of days of the week on accidents. The weekly pattern of accidents is shown in Figure 4. The result revealed that there was not a statistically significant difference in accidents on weekdays ($p > 0.05$). The result is shown in Table 1.

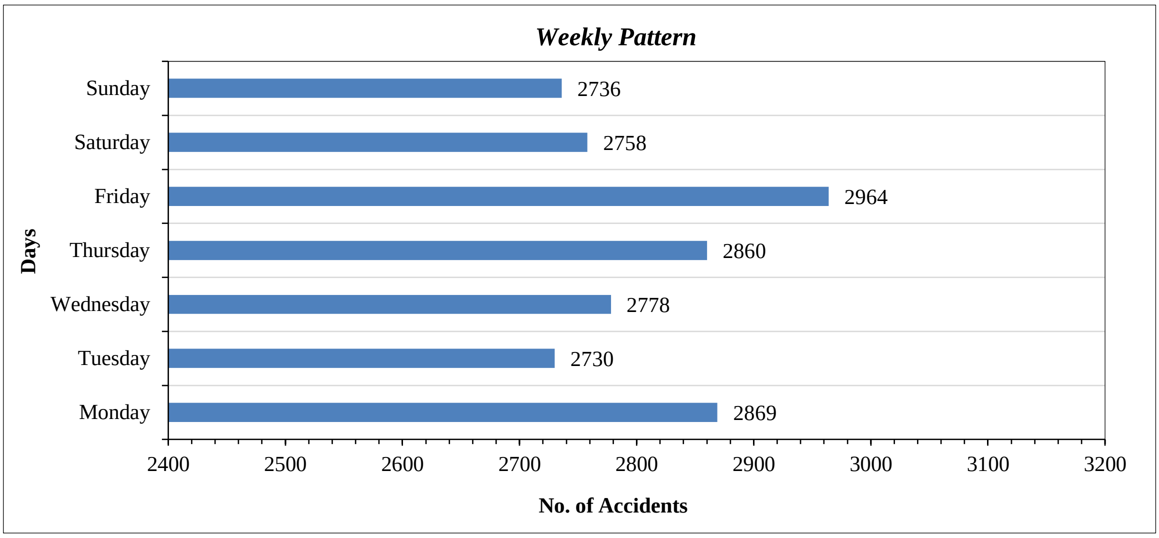

Figure 4. No. of accidents throughout the weeks of the year 2022

Table 1. One Way ANOVA test results for weekdays and accident variables

	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	41.779	6	6.963	1.698	0.117
Within Groups	35705.220	8705	4.102		
Total	35746.999	8711			

To investigate the effects of time on accident occurrence, One Way ANOVA test was applied in SPSS. The temporal pattern of the accidents is shown in Figure 5. The result showed that there was a statistically significant difference in accidents occurring at different times of the day ($p < 0.001$). The result is shown in Table 2.

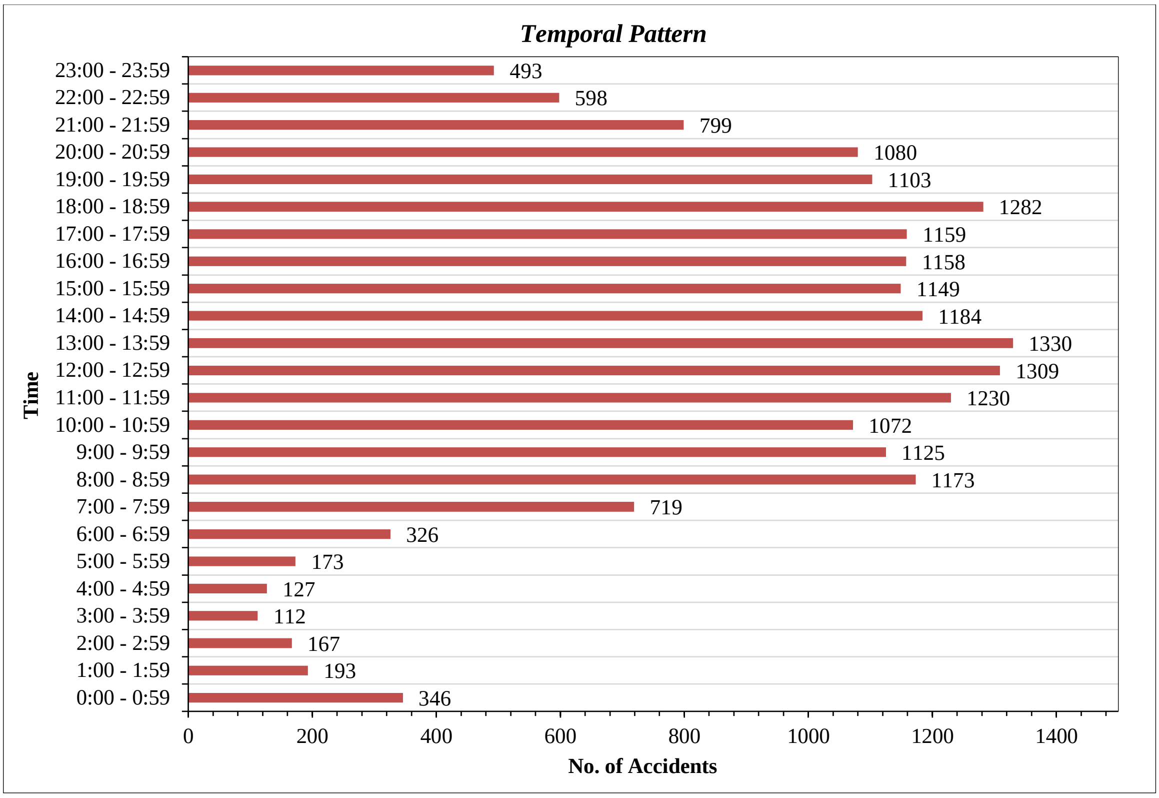

Figure 5. Accidents at different times of the day throughout the year 2022

Table 2. One Way ANOVA test results for time and accident variables

	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	13162.101	23	572.265	220.140	0.000
Within Groups	22584.898	8688	2.600		
Total	35746.999	8711			

3.1.1. Seasonal Patterns of Accidents

Pakistan is located in the northern hemisphere, and throughout the year, it has four seasons lasting every three months. The spring season includes the months of March, April, and May. The summer season occurs during June, July, and August. The autumn season consists of September, October, and November, while the winter season occurs during December, January, and February [57]. In spring, there were 5642 accidents, the highest among all the seasons. One of the reasons might include pleasant weather during spring, which encourages more people to go outside. The increased outdoor activity causes more accidents during this season. This pattern is shown in Figure 6.

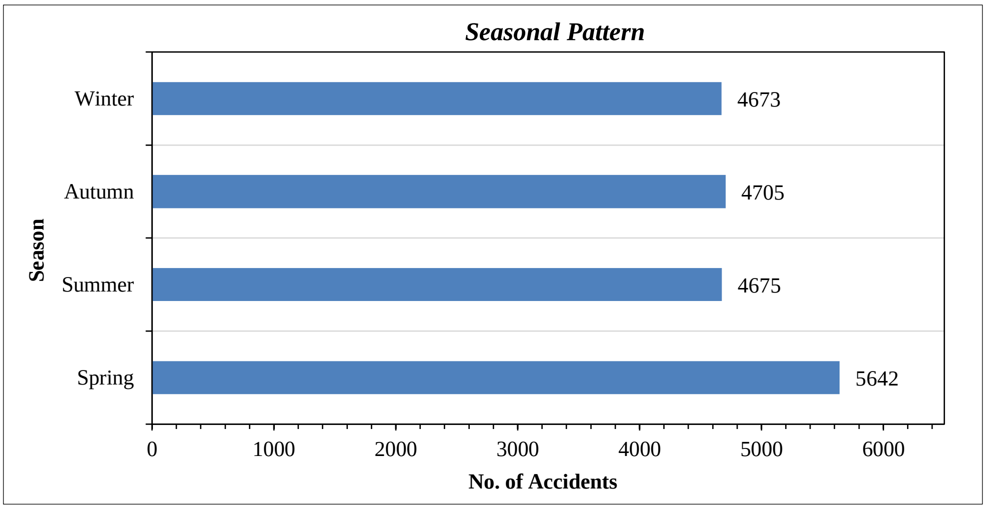

Figure 6. Accidents in different Seasons of the year 2022

3.1.2. Monthly Pattern of Accidents

The pattern of accidents during each month of the year was found to be random. The lowest number of accident spots occurred in January, i.e., 1448 which could be linked to colder weather and potentially fewer people on the roads. The pattern is shown in Figure 7.

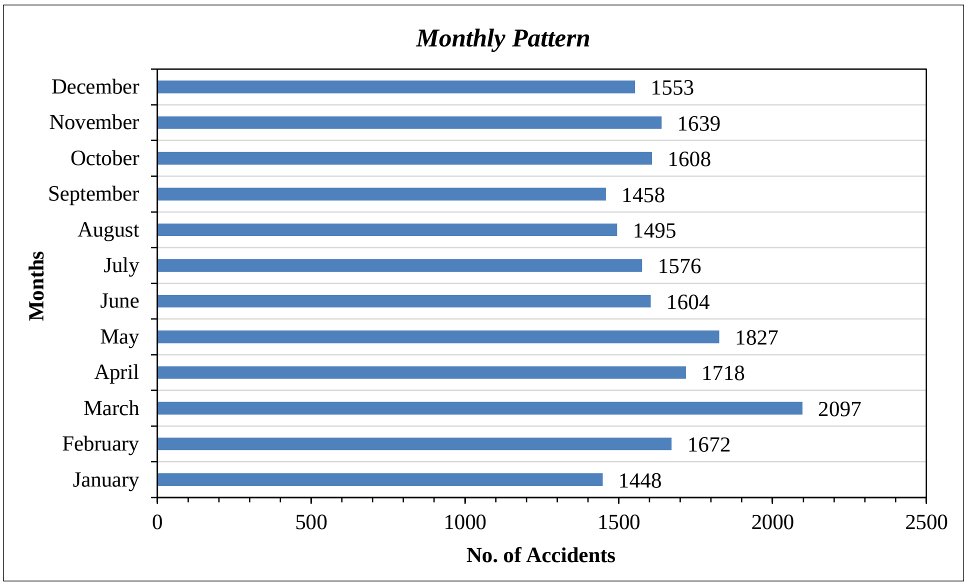

Figure 7. Accidents in different months of the year 2022

3.1.3. Weekly Pattern of Accidents

There is high traffic volume on the roads during the working days of the week. This leads to congestion and aggressive driver behavior. Hence, weekdays experience more accidents in an urban area with high commercial activities. There was a slightly increased pattern of accidents on Friday. As Faisalabad is an industrial city, workers leave for their houses on Friday (the last working day of the week). Mondays often experience more accidents, potentially due to driver weariness and a busy start to a new week of schools and offices. Mondays and Fridays face the highest number of traffic accidents; the number of accident spots on Mondays was 2869, while on Fridays, it was 2964 throughout the year 2022. The pattern is shown in Figure 4.

3.1.4. Temporal Distribution of Traffic Accident Hotspots

The temporal distribution of traffic accident hotspots in Faisalabad, Pakistan, varies daily. Peak hours, including morning (08 a.m.–11 a.m.), afternoon (11 p.m.–02 p.m.), and evening (03 p.m.–07 p.m.) rush hours, experience high traffic volumes and increased accident risks.

As shown in Figure 7, the highest number of traffic accidents (1330) occurred during the afternoon (01 p.m.–02 p.m.). This was authenticated by One Way ANOVA test on the accidents occurring at different times of the day. The result is shown in Table 2. This specific pattern was found since the closing of schools caused the roads to block or traffic jams, increasing the chances of an accident. Also, during the evening, rush hours increase due to the closing of offices. While in the morning, the chances of accidents elevate due to the opening of schools and offices. At night, accidents pose unique challenges due to reduced visibility and potential driver fatigue. Transition times, such as dawn and dusk, require extra caution due to changing visibility conditions. Specific factors, like school zones and increased pedestrian activity, also influence accident patterns.

3.2. Results of ArcGIS

The local Moran’s I static was applied to generate a significant map of clusters with statistical meaning. The result of Moran's I Static indicates the hotspots and cold spots of road traffic data, as shown in Figure 8. It shows that there are 140 hotspots, out of which 117 hotspots are of 99% confidence level, 16 hotspots are of 95% confidence level, and 7 hotspots are of 90% confidence level. Most of the hotspots are located around the main GTS stop because of the high volume of vehicles in these areas and the presence of major bus stations in Faisalabad. Most of the labor workers leave the city for work, and the pedestrian rate is higher near the bus stands. The study investigated traffic accident hotspots in Faisalabad's urban area using traffic accident data for the year 2022, uncovering key areas of concern and contributing to road safety strategies. Using the Local Moran's I approach, findings revealed 140 hotspots, mainly clustered around a 4 km radius from the GTS Stop, and 62 coldspots in the outskirts and boundaries of the study area. This area consists of schools, markets, hospitals, and bus stops. Hence, excessive business activities occur in this area during the daytime. The increased number of road users causes an increase in traffic accidents during the daytime. The maximum number of accidents were found during 01–02 pm. During this peak hour, traffic flow is at its maximum, causing an increase in accidents. The study can help policymakers in selecting different measures at specific locations to reduce the number of accidents and improve traffic safety.

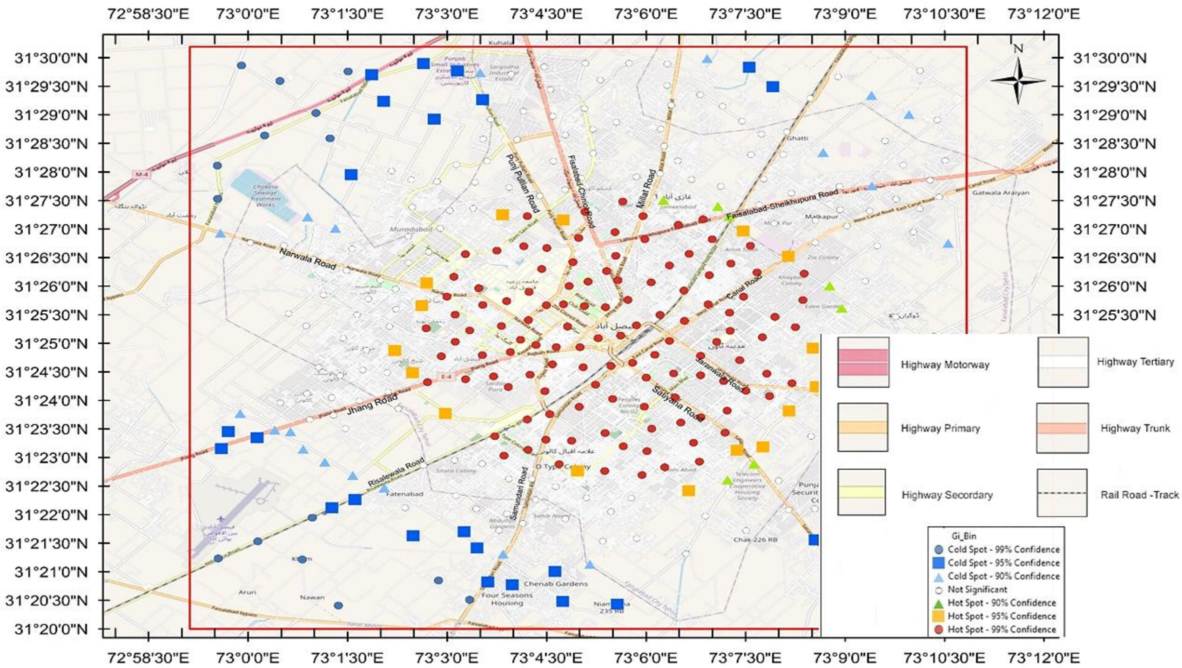

Figure 8. Hotspots & Cold spots of traffic accidents in the year 2022

3.3. Results of Contributing Factors

While developing the traffic accident prediction model, attributes showed different kinds of results according to their effectiveness in the model. The result is shown in Table 3. The positive correlation implies that as these attributes experience certain changes or fluctuations, the likelihood of a traffic accident also increases. The analysis of road measurements has revealed a notable and affirmative correlation with the frequency of accidents. The correlation results of different contributing factors showed that the nature of traffic and road dimensions were the attributes that showed a positive correlation with the occurrence of accidents. In the urban area of Faisalabad, mixed traffic consists of vulnerable road users, passengers, heavy vehicles, etc. Due to this heterogeneous mixture of traffic, the probability of an accident occurring increases. This nature of traffic also added to accidents. Because of increased urban traffic, the existing road infrastructure contains various deficiencies in design. Lack of designated lanes, segregation of vulnerable road users, and an increased number of conflict points are a few of the prominent issues related to road dimensions.

Table 3. Co-relation Results

Co-Attributes	Co-relations
Motor vehicle incidents	1
Road measurements	0.679894
Heavy traffic	0.361761
Diversion	0.339741
Disconnector number of traffic lights	0.240451
Road diversion	0.205066
Age factor	-0.364596
Number of turns	-0.374693
Green belt width	-0.393715
Number of pedestrian crossing	-0.422873
Type of vehicle	-0.456657
In combating traffic	-0.473516
Public building school/college	-0.493534

3.4. Results of Accident Prediction

From the applied machine learning algorithms, random forest showed the minimum variation with the actual values. However, linear regression and decision trees also showed good accuracy, but less than random forest. The Root Mean Square Error makes the difference because its lower value means more precision. All the applied machine learning algorithms had an accuracy of a minimum of 75%. From Table 4, among the three models used for accident prediction, the Decision Tree model was found to be the most accurate, with an accuracy of 84%. All three machine learning algorithms gave good accuracy in accident prediction. The results pattern resembled past research work conducted with similar attributes and models [46]. To minimize traffic accidents, speed restrictions and effective law enforcement by the agencies can be crucial in reducing accident numbers. To curtail the traffic vehicles in the hotspot areas, alternative routes for the through traffic might reduce the accident occurrence in the area.

Table 4. Results of Machine Learning Models

Algorithms	Root Mean Square Error (RMSE)	Standard Deviation	Accuracy (%)
Random Forest	2.94	0.69	76.2
Decision Tree	3.9	0.83	84.4
Linear Regression	4.14	1.06	75

4. Conclusion

This research was conducted to predict traffic accidents in an urban area. To achieve this goal, traffic accident hotspots were identified using ArcGIS. Furthermore, the pattern of accidents and the effects of contributing factors were also analyzed. A large number of accident hotspots were identified all around the urban road network of Faisalabad using ArcGIS. This accident concentration was highest around the General Transport Stand in Faisalabad. This implies that remedial measures should be applied in this particular zone to reduce the accident frequency. Furthermore, the accident number was highest during the daytime from 01 p.m. to 02 p.m. This shows that traffic flow during this hour is at its maximum in the whole urban area, and it is not being managed effectively. Thus, strong remedial management techniques are needed to decrease the number of accidents during this time.

Analysis of the correlation between contributing factors and traffic accidents revealed that Road Measurements have a significant effect on accident occurrence. As road width in densely populated areas is difficult to increase, effective alternatives should be provided by the authorities to minimize traffic accidents. Furthermore, the separation of vulnerable road users, where possible, is highly recommended.

Accident prediction using developed machine algorithms can help assess the danger of accident-prone locations before accidents occur. This will enable authorities to take effective actions to minimize accidents.

In the future, researchers can ensure more studies using different techniques, like using global techniques to find dangerous spots, and can compare the results with this research. The role of the geometric features of the roads at these hotspots can also be investigated. The police and health departments should also have the traffic crash information so that the data can be verified from multiple sources. Data is the key to any aspect of research. Hence, the relevant departments must provide an increased range of data in the future for a better understanding of past patterns and accuracy. A large amount of input data points must be used in the future for better accuracy of machine learning models.

4.1. Limitations

The Rescue 1122 Emergency Services Department supplied only one year of traffic accident data. While this data is insightful, a more extensive dataset spanning a longer period of time could offer a more refined understanding of trends and patterns. Moreover, the coordinates provided in our research dataset corresponded to the locations of ID callers rather than pinpointing the exact locations of the actual traffic accident. This aspect might introduce a level of imprecision in our analysis. The lack of information on the type of crash, weather conditions at the time of accidents' occurrence, etc. potentially limits a thorough exploration of contributing factors and potential correlations. This also limits the accuracy of the prediction models.

5. Declarations

5.1. Author Contributions

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

5.2. Data Availability Statement

The data presented in this study are available in the article.

5.3. Funding

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

5.4. Acknowledgements

The authors thank Rescue 1122 Emergency Services Faisalabad for providing the very rich accident dataset. The authors also thank the City Traffic Police Faisalabad for facilitating the Field Survey for the collection of traffic and geometric data on roads.

5.5. Conflicts of Interest

The authors declare no conflict of interest.

6. References

[1] WHO. (2022). Road traffic injuries. World Health Organization (WHO), Geneva, Switzerland. Available online: https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries (accessed on May 2024).

[2] PBS. (2023). Announcement of Results of 7th Population and Housing Census-2023 'The Digital Census'. Pakistan Bureau of Statistics, Islamabad, Pakistan. Available online: https://www.pbs.gov.pk/content/announcement-results-7th-population-and-housing-census-2023-digital-census (accessed on March 2024).

[3] World Life Expectancy. (2023). World health rankings. Available online: https://www.worldlifeexpectancy.com/cause-of-death/road-traffic-accidents/by-country/ (accessed on May 2024).

[4] Rais, R. B. (2023). Road safety in Pakistan is the not-so-silent killer everybody is ignoring. ARAB NEWS, Riyadh, Saudi Arabia.

[5] Pusuluri, V. L., & Dangeti, M. R. (2024). Applications of QGIS and machine learning for road crash spot identification. Earth Science Informatics, 17(3), 2331–2346. doi:10.1007/s12145-024-01271-0.

[6] Khatun, M. S., Hossain, M. A., Kabir, M. A., & Rahman, M. A. (2024). Identification and analysis of accident black spots using Geographic Information System (GIS): A study on Kushtia-Jhenaidah national highway (N704), Bangladesh. Heliyon, 10(3), e25952. doi:10.1016/j.heliyon.2024.e25952.

[7] Chun, U., Lim, J., & Kim, H. (2024). Analysis of Hotspots in and outside School Zones: A Case Study of Seoul. Journal of Advanced Transportation, 2024, 1–13. doi:10.1155/2024/6613603.

[8] Hammas, M., & Al-Modayan, A. (2022). Spatial Analysis of Traffic Accidents in the City of Medina Using GIS. Journal of Geographic Information System, 14(05), 462–477. doi:10.4236/jgis.2022.145025.

[9] Feizizadeh, B., Omarzadeh, D., Sharifi, A., Rahmani, A., Lakes, T., & Blaschke, T. (2022). A GIS-Based Spatiotemporal Modelling of Urban Traffic Accidents in Tabriz City during the COVID-19 Pandemic. Sustainability (Switzerland), 14(12), 7468. doi:10.3390/su14127468.

[10] Rahman, M. T., Jamal, A., & Al-Ahmadi, H. M. (2020). Examining hotspots of traffic collisions and their spatial relationships with land use: A GIS-based geographically weighted regression approach for Dammam, Saudi Arabia. ISPRS International Journal of Geo-Information, 9(9), 540. doi:10.3390/ijgi9090540.

[11] ESRI. (2024). Geographic Information System. Geographic information system company (ESRI), Redlands, United States. Available online: https://www.esri.com/en-us/what-is-gis/overview (accessed on May 2024).

[12] Le, K. G., Liu, P., & Lin, L.-T. (2020). Traffic accident hotspot identification by integrating kernel density estimation and spatial autocorrelation analysis: a case study. International Journal of Crashworthiness, 27(2), 543–553. doi:10.1080/13588265.2020.1826800.

[13] Zahran, E.-S. M. M., Tan, S. J., Tan, E. H. A., Mohamad 'Asri Putra, N. A. 'Atiqah B., Yap, Y. H., & Abdul Rahman, E. K. (2019). Spatial analysis of road traffic accident hotspots: evaluation and validation of recent approaches using road safety audit. Journal of Transportation Safety & Security, 13(6), 575–604. doi:10.1080/19439962.2019.1658673.

[14] Al-Aamri, A. K., Hornby, G., Zhang, L.-C., Al-Maniri, A. A., & Padmadas, S. S. (2021). Mapping road traffic crash hotspots using GIS-based methods: A case study of Muscat Governorate in the Sultanate of Oman. Spatial Statistics, 42, 100458. doi:10.1016/j.spasta.2020.100458.

[15] Srikanth, Dr. L., Srikanth, I., & Arockiasamy, Dr. M. (2019). Identification of Traffic Accident Hotspots using Geographical Information System (GIS). International Journal of Engineering and Advanced Technology, 9(2), 4429–4438. doi:10.35940/ijeat.b3848.129219.

[16] Afolayan, A., Easa, S. M., Abiola, O. S., Alayaki, F. M., & Folorunso, O. (2022). GIS-Based Spatial Analysis of Accident Hotspots: A Nigerian Case Study. Infrastructures, 7(8), 103. doi:10.3390/infrastructures7080103.

[17] Kričković, E., Lukić, T., Srejić, T., Stojšić-Milosavljević, A., Stojanović, V., & Kričković, Z. (2024). Spatial-temporal and trend analysis of traffic accidents in AP Vojvodina (North Serbia). Open Geosciences, 16(1), 20220630. doi:10.1515/geo-2022-0630.

[18] Umair, M., Rana, I. A., & Lodhi, R. H. (2022). The impact of urban design and the built environment on road traffic crashes: A case study of Rawalpindi, Pakistan. Case Studies on Transport Policy, 10(1), 417–426. doi:10.1016/j.cstp.2022.01.002.

[19] Maurya, R. K., & Swain, J. B. (2022). Modeling and Analysis of Accident in an Urban Area (No. 9009). EasyChair, Wythenshawe, United Kingdom.

[20] Srikanth, L., & Srikanth, I. (2020). A Case Study on Kernel Density Estimation and Hotspot Analysis Methods in Traffic Safety Management. 2020 International Conference on COMmunication Systems & NETworkS (COMSNETS), Bengaluru, India. doi:10.1109/comsnets48256.2020.9027448.

[21] Ahmad, A., Hossain, K., & Hossain, M. (2020). Identification of Urban Traffic Accident Hotspot Zones Using GIS: A Case Study of Dhaka Metropolitan Area. Journal of Geographical Studies, 3(1), 36–42. doi:10.21523/gcj5.19030104.

[22] Alam, M. S., & Tabassum, N. J. (2023). Spatial pattern identification and crash severity analysis of road traffic crash hot spots in Ohio. Heliyon, 9(5), e16303. doi:10.1016/j.heliyon.2023.e16303.

[23] Mekonnen, A. A., Sipos, T., & Krizsik, N. (2023). Identifying Hazardous Crash Locations Using Empirical Bayes and Spatial Autocorrelation. ISPRS International Journal of Geo-Information, 12(3), 85. doi:10.3390/ijgi12030085.

[24] Lee, A. S., Lin, W. H., Gill, G. S., & Cheng, W. (2019). An enhanced empirical bayesian method for identifying road hotspots and predicting number of crashes. Journal of Transportation Safety & Security, 11(5), 562-578. doi:10.1080/19439962.2018.1450314.

[25] arirforoush, H., Bellalite, L., & Bénié, G. B. (2019). Spatial and Temporal Analysis of Seasonal Traffic Accidents. American Journal of Traffic and Transportation Engineering, 4(1), 7. doi:10.11648/j.ajtte.20190401.12.

[26] Hovenden, E., & Liu, G.-J. (2020). Use of Spatial Analysis Techniques to Identify Statistically Significant Crash Hot Spots in Metropolitan Melbourne. Journal of Road Safety, 31(4), 36–58. doi:10.33492/jrs-d-19-00249.

[27] Qalb, A., Arshad, H. S. H., Nawaz, M. S., & Hafeez, A. (2023). Risk reduction via spatial and temporal visualization of road accidents: a way forward for emergency response optimization in developing countries. International Journal of Injury Control and Safety Promotion, 30(2), 310–320. doi:10.1080/17457300.2022.2164312.

[28] Erenler, A. K., & Gümüş, B. (2019). Analysis of road traffic accidents in Turkey between 2013 and 2017. Medicina (Lithuania), 55(10), 679. doi:10.3390/medicina55100679.

[29] Alkhadour, W., Zraqou, J., Al-Helali, A., & Al-Ghananeem, S. (2021). Traffic Accidents Detection using Geographic Information Systems (GIS). International Journal of Advanced Computer Science and Applications, 12(4), 484-494. doi:10.14569/ijacsa.2021.0120462.

[30] Hazaymeh, K., Almagbile, A., & Alomari, A. H. (2022). Spatiotemporal Analysis of Traffic Accidents Hotspots Based on Geospatial Techniques. ISPRS International Journal of Geo-Information, 11(4), 260. doi:10.3390/ijgi11040260.

[31] Wang, Z. Z., Lu, Y. N., Zou, Z. H., Ma, Y. H., & Wang, T. (2022). Applying OHSA to Detect Road Accident Blackspots. International Journal of Environmental Research and Public Health, 19(24), 16970. doi:10.3390/ijerph192416970.

[32] Unsar, S., & Sut, N. (2009). General assessment of the occupational accidents that occurred in Turkey between the years 2000 and 2005. Safety Science, 47(5), 614-619. doi:10.1016/j.ssci.2008.08.001.

[33] Le, K. G., Liu, P., & Lin, L. T. (2020). Determining the road traffic accident hotspots using GIS-based temporal-spatial statistical analytic techniques in Hanoi, Vietnam. Geo-spatial Information Science, 23(2), 153-164. doi:10.1080/10095020.2019.1683437.

[34] Vyshnavi, K. G., & Nalini, V. K. (2022). Machine Learning Algorithms for Road Accident Analysis and Forecasting. International Journal of Research in Engineering and Science, 10, 283–288.

[35] Venkat, A., Vijey, G. K., & Susan Thomas, I. (2020). Machine Learning Based Analysis for Road Accident Prediction. International Journal of Emerging Technology and Innovative Engineering, 6(2), 2394–6598.

[36] Megibaru, L., & Atnafu, B. (2019). Design a Machine Learning Model to Identify and Predict the Cause of Vehicle Accidents in East Gojjam, Amhara, Ethiopia. International Journal of Innovative Research in Science, Engineering and Technology, 8(12), 11856–11863.

[37] Guerra, A., Gadhiya, V., & Srisurin, P. (2022). Crash Prediction on Road Segments Using Machine Learning Methods. ASEAN Engineering Journal, 12(3), 27–37. doi:10.11113/AEJ.V12.17601.

[38] Dogru, N., & Subasi, A. (2018). Traffic accident detection using random forest classifier. 2018 15th Learning and Technology Conference (L&T). doi:10.1109/lt.2018.8368509.

[39] Sripuram, A., Jayasree., T., Vaishnavi, T., Reddy, V. R., & Mtech, P. C. (2022). Road Accident Analysis using Machine Learning. International Research Journal of Modernization in Engineering Technology and Science, 4(3), 1364-1371.

[40] Ulu, M., Kilic, E., & Türkan, Y. S. (2024). Prediction of Traffic Incident Locations with a Geohash-Based Model Using Machine Learning Algorithms. Applied Sciences, 14(2), 725. doi:10.3390/app14020725.

[41] Sridevi, N., Keerthana, M. V., Pal, M. V., Nikshitha, T. R., & Jyothi, P. (2020). Road accident analysis using machine learning. International Journal of Research in Engineering, Science and Management, 3(5), 859-861.

[42] Azimjonov, J., Özmen, A., & Varan, M. (2023). A vision-based real-time traffic flow monitoring system for road intersections. Multimedia Tools and Applications, 82(16), 25155–25174. doi:10.1007/s11042-023-14418-w.

[43] Torres, K. A. R., & Asor, J. R. (2021). Machine learning approach on road accidents analysis in Calabarzon, Philippines: An input to road safety management. Indonesian Journal of Electrical Engineering and Computer Science, 24(2), 993–1000. doi:10.11591/ijeecs.v24.i2.pp993-1000.

[44] Růžička, J., Bělinová, Z., Korec, V., Cikhardtová, K., & Tichý, T. (2019). Prediction of traffic accidents by using neural network. Archives of Transport System Telematics, 12(4), 22-26.

[45] Jadhav, A., Jadhav, S., Jalke, A., & Suryavanshi, K. (2020). Road accident analysis and prediction of accident severity using machine learning. International Research Journal of Engineering and Technology (IRJET), 7(12), 740-747.

[46] Santos, D., Saias, J., Quaresma, P., & Nogueira, V. B. (2021). Machine learning approaches to traffic accident analysis and hotspot prediction. Computers, 10(12), 157. doi:10.3390/computers10120157.

[47] Candefjord, S., Muhammad, A. S., Bangalore, P., & Buendia, R. (2021). On Scene Injury Severity Prediction (OSISP) machine learning algorithms for motor vehicle crash occupants in US. Journal of Transport and Health, 22. doi:10.1016/j.jth.2021.101124.

[48] Najafi Moghaddam Gilani, V., Hosseinian, S. M., Ghasedi, M., & Nikookar, M. (2021). Data-Driven Urban Traffic Accident Analysis and Prediction Using Logit and Machine Learning-Based Pattern Recognition Models. Mathematical Problems in Engineering, 2021, 1–11. doi:10.1155/2021/9974219.

[49] Yassin, S. S., & Pooja. (2020). Road accident prediction and model interpretation using a hybrid K-means and random forest algorithm approach. SN Applied Sciences, 2(9). doi:10.1007/s42452-020-3125-1.

[50] Rana, V., Joshi, H., Parmar, D., Jadhav, P., & Kanojiya, M. (2019). Road accident prediction using machine learning algorithm. International Research Journal of Engineering and Technology, 6(3), 3934-3496.

[51] ESRI. (2023). ArcGIS: Geographic Information System Company (ESRI), Redlands, United States. Available online: https://www.esri.com/en-us/arcgis/about-arcgis/overview (accessed on March 2024).

[52] Harirforoush, H., & Bellalite, L. (2019). A new integrated GIS-based analysis to detect hotspots: A case study of the city of Sherbrooke. Accident Analysis and Prevention, 130, 62–74. doi:10.1016/j.aap.2016.08.015.

[53] Aghasi, N. H. M. (2019). Application of GIS for Urban Traffic Accidents: A Critical Review. Journal of Geographic Information System, 11(01), 82–96. doi:10.4236/jgis.2019.111007.

[54] Gedamu, W. T., Plank-Wiedenbeck, U., & Wodajo, B. T. (2024). A spatial autocorrelation analysis of road traffic crash by severity using Moran's I spatial statistics: A comparative study of Addis Ababa and Berlin cities. Accident Analysis & Prevention, 200, 107535. doi:10.1016/j.aap.2024.107535.

[55] Manap, N., Borhan, M. N., Yazid, M. R. M., Hambali, M. K. A., & Rohan, A. (2021). Identification of Hotspot Segments with a Risk of Heavy-Vehicle Accidents Based on Spatial Analysis at Controlled-Access Highway. Sustainability, 13(3), 1487. doi:10.3390/su13031487.

[56] Haybat, H., Zerenoğlu, H., & Özlü, T. (2022). Temporal and Spatial Analysis of Traffic Accidents: The Case of Bursa City. International Journal of Geography and Geography Education, 45, 404–423. doi:10.32003/igge.1016204.

[57] Seasons of the Year. (2024). Seasons in Pakistan. Seasons of the Year, Malakhovka, Russia. Available online: https://seasonsyear.com/Pakistan (accessed on March 2024).