Traffic Crash Analysis of Three-Wheelers in Rawalpindi: Prediction Modeling using Machine Learning and Emerging Hotspot Analysis

By

Abdul Moeed

(Registration No: 00000361953)

Department of Transportation Engineering

School of Civil and Environmental Engineering

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

(2024)

Traffic Crash Analysis of Three-Wheelers in Rawalpindi: Prediction Modeling using Machine Learning and Emerging Hotspot Analysis

By

Abdul Moeed

(Registration No: 00000361953)

A thesis submitted to the National University of Sciences and Technology, Islamabad,

in partial fulfillment of the requirements for the degree of

Master of Science in
Transportation Engineering

Supervisor: Dr. Muhammad Asif Khan

Co Supervisor: Dr. Kamran Ahmed

School of Civil and Environmental Engineering

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

(2024)

THESIS ACCEPTANCE CERTIFICATE

It is certified that final copy of MS thesis written by **Mr. Abdul Moeed**, Registration No. **00000361953**, of MS Transportation Engineering SCEE, (NICE), has been vetted by undersigned, found completed in all respects as per NUST Statutes/Regulations, is free of plagiarism, errors, and mistakes and is accepted as partial fulfilment for award of MS degree.

Signature: ____________________

Supervisor: Dr. Muhammad Asif Khan

Date: 01/10/2024

Dr. Arshad Hussain
Associate Professor
Head of Transportation Engineering Department
School of Civil & Environmental Engineering
NUST, Sector H-12, Islamabad

Signature: ____________________

Head of Department: Dr. Arshad Hussain

Date: 01-10-2024

Dr. S. Muhammad Jamil
Associate Dean
NICE, SCEE, NUST

Signature: ____________________

Associate Dean (NICE): Dr. S. Muhammad Jamil

Date: 08-10-2024

Signature: ____________________

Principal & Dean (SCEE): Prof. Dr. Muhammad Irfan

Date: 09 OCT 2024

PROF DR MUHAMMAD IRFAN
Principal & Dean
SCEE, NUST

Form TH-4

National University of Sciences and Technology

MASTER'S THESIS WORK

We hereby recommend that the dissertation prepared under our Supervision by: Abdul Moeed, Regn No. 00000361953

Titled: **Traffic Crash Analysis of Three-Wheelers in Rawalpindi: Prediction Modeling using Machine Learning and Emerging Hotspot Analysis**

be accepted in partial fulfillment of the requirements for the award of **Master of Science** degree with (B+ Grade).

Examination Committee Members

1. Name: Dr. Arshad Hussain — Signature: ____________

2. Name: Dr. Sameer Ud Din — Signature: ____________

Co Supervisor's Name: Dr. Kamran Ahmed — Signature: ____________

Supervisor's Name: Dr. Muhammad Asif Khan — Signature: ____________

Date: 27/8/2024

Head of Department
Dated: 08/10/2024

Associate Dean
Dated 08/10/2024
Dr. S. Muhammad Jamil
Associate Dean
NICE, SCEE, NUST

Associate Dean (NICE
Dated: 08/10/2024

COUNTERSIGNED

Principal & Dean SCEE
PROF DR MUHAMMAD IRFAN
Principal & Dean
SCEE, NUST

Date: 09 OCT 2024

Principal & Dean SCEE

Certificate of Approval

This is to certify that the research work presented in this thesis, entitled Traffic Crash Analysis of Three-Wheelers in Rawalpindi: Prediction Modeling using Machine Learning and Emerging Hotspot Analysis**"** was conducted by Mr. Abdul Moeed under the supervision of Dr. Muhammad Asif Khan.

No part of this thesis has been submitted anywhere else for any other degree. This thesis is submitted to the Department of Transportation Engineering in partial fulfillment of the requirements for the Degree of Masters of Science in Field of Transportation Engineering from NUST Institute of Civil Engineering (NICE), School of Civil and Environmental Engineering (SCEE), NUST.

Student Name: Abdul Moeed — Signature:

Examination Committee

a) GEC Member 1. Dr. Arshad Hussain — Signature

Associate Professor

b) GEC Member 2. Dr. Sameer Ud Din — Signature

Assistant Professor

Co Supervisor Name: Dr. Kamran Ahmed — Signature

Supervisor Name: Dr. Muhammad Asif Khan — Signature

Name of HoD: Dr. Arshad Hussain — Signature

HoD Transportation Engineering
NUST Institute of Civil Engineering
School of Civil & Environmental Engineering
National University of Sciences and Technology

Name of Associate Dean: Dr. S. Muhammad Jamil — Signature

Dr. S. Muhammad Jamil
Associate Dean
NICE, SCEE, NUST

Name of Principal & Dean: Dr. Muhammad Irfan — Signature

PROF DR MUHAMMAD IRFAN
Principal & Dean
SCEE, NUST

AUTHOR'S DECLARATION

I Abdul Moeed hereby state that my MS thesis titled "Traffic Crash Analysis of Three-Wheelers in Rawalpindi: Prediction Modelling using Machine Learning and Emerging Hotspot Analysis" is my own work and has not been submitted previously by me for taking any degree from National University of Sciences and Technology, Islamabad or anywhere else in the country/ world.

At any time if my statement is found to be incorrect even after I graduate, the university has the right to withdraw my MS degree.

Name of Student: Abdul Moeed

Date: October 9, 2024

PLAGIARISM UNDERTAKING

I solemnly declare that research work presented in the thesis titled "Traffic Crash Analysis of Three-Wheelers in Rawalpindi: Prediction Modelling using Machine Learning and Emerging Hotspot Analysis" is solely my research work with no significant contribution from any other person. Small contribution/ help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero-tolerance policy of the HEC and National University of Science and Technology towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS degree, the University reserves the rights to withdraw/revoke my MS degree and that HEC and the University has the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized thesis.

Signature:________________________________

Name of Student: Abdul Moeed

Dedicated to my Parents,

Siblings and

Teachers

ACKNOWLEDGEMENTS

All praise be to ALLAH, I am grateful to Almighty ALLAH, Who made me capable of achieving this milestone in my life with his countless blessings. Secondly, I would like to thank my loving parents and my brother Arsalan Shaukat without their prayers and support I would not be able to complete this research, their cooperation and understanding has played a huge role in this achievement. I am deeply appreciative of my research supervisor Dr. Muhammad Asif, whose guidance, innovative insights, and invaluable feedback were instrumental in helping me achieve this goal. I am also thankful to my co-supervisor Dr. Kamran Ahmed who made sure to make my research hassle free through his support in process of data acquisition and data handling. I would also like to thank my GEC members Dr. Arshad Hussain and Dr. Sameer Ud Din for their valuable guidance throughout my Research journey.

I convey my sincere appreciation to Rescue 1122 department Rawalpindi and Regional Emergency Officer Dr. Abdur Rahman for providing me traffic crash data and Pakistan Meteorological Department (PMD) for sharing relevant climate data. Their cooperation throughout this data acquisition process is highly appreciated.

TABLE OF CONTENTS

LIST OF TABLES

Page No.

LIST OF FIGURES

Page No.

LIST OF SYMBOLS, ABBREVIATIONS AND ACRONYMS

3-MR	Three Wheeled Motorized Rickshaw
ANN	Artificial Neural Network
BRT	Bus Rapid Transit
GIS	Geographic Information System
ISA	Incremental Spatial Analysis
ROC	Receiver Operating Curve
RTC	Road Traffic Crashes
SMOTE	Synthetic Minority Oversampling Technique
WHO	World Health Organization

ABSTRACT

The three-wheeled motorized rickshaw (3-MR) is the dominant mode of transportation in developing countries for short trips with passenger carrying capacity of four to six and movement of goods at a small-scale level. These 3-MRs are often associated with road traffic crashes causing significant socioeconomic and public health concerns. Early research mostly focused on the safety analysis of two and four-wheelers, leading to a research gap in exploring the safety dynamics associated with 3-MR crashes. Few recent studies investigated the effect of different factors on injury severity of 3-MR crashes using statistical and machine learning models, but no study evaluated the spatiotemporal dimension of the 3-MR crashes. This study consists of two parts. In first part study aims to identify crash hotspots using 3-MR crash data for sixteen months (January 2022 to April 2023) in Rawalpindi, Pakistan. The study employs cutting-edge spatial analytic methods, such as emerging hotspot analysis with a space-time cube and global and local spatial autocorrelation. Significant clustering in the form of hotspots is observed, employing Moran's index for global spatial autocorrelation detection, and local spatial autocorrelation was assessed using Getis Ord Gi*, which supports a recurring pattern in the 3-MR crashes. Incremental spatial autocorrelation was also employed, indicating 272 meters as the optimal distance for clustering detection. Further utilizing the emerging hotspot analysis technique, substantial spatiotemporal clustering is identified. Critical hotspots were identified near major bus stops, commercial areas, intersections, hospitals, airports, and high-density residential units. In second part Artificial neural network model with back propagation algorithm was developed to predict injury severity of 3-MR traffic crashes using some key variables like emergency response time, that were not considered in past research. Synthetic Minority Oversampling Technique (SMOTE) was also employed to deal with class imbalance between Severe and Minor injury severity. Model achieved an overall accuracy of 74.5% and Precision of 75.1% with true positive ratio of 73.7% and false positive ratio of 75.3%. Receiver Operating Characteristics (ROC) curve was also analyzed, area under the curve value of 0.809 for both severe and minor injury classes depicted ability of model to distinguish between both classes. Sensitivity Index further ranked predictors according to variable importance and their impact on injury severity. Model identified higher response times, over speeding, distracted driving behaviors, and higher temperature increases the risk of severe injury crash. The findings of our study have significant practical implications for stakeholders, providing them with a foundation and pinpoint locations in the form of hotspots and key factors contributing to crash injury severity, to update and refine policies specifically aimed at addressing recurrent 3-MR crashes and enhancing overall road traffic safety for vulnerable road users.

Keywords: Emerging Hotspot Analysis, Space-Time Cube, Three-wheeler Motorized Rickshaw Crashes, Global Spatial Autocorrelation, Incremental Spatial Autocorrelation, Local Spatial Autocorrelation

CHAPTER 1: INTRODUCTION

1.1 Background

Road traffic injuries are one of the leading causes of death across the world for individuals in the age group of 15 to 29 (World Health Organization, 2018). Statistically, every year road crashes lead to 1.35 million fatalities worldwide. Southeast Asia leads the world in the number of road deaths at 20.7 per 100,000 people, compared to 15.6 and 9.3 in America and Europe, respectively. Further, it is evident that a higher percentage of traffic-related deaths take place in nations with lower and moderate levels of resources. Developing nations demonstrate higher rates of road traffic injuries, with 93% of fatalities originating from low- and middle-income countries (World Health Organization, 2015). Vulnerable road users, such as pedestrians, cyclists, and motorcyclists, account for more than 50% of all fatalities and injuries in road traffic events (World Health Organization, 2018). About 286000 fatalities occur due to motorized two and 3-MR, which accounts for 23% of total road traffic crash (RTC) fatalities (World Health Organization, 2017).

Urban transportation networks are crucial in determining how cities function and how citizens' quality of life is affected. Punjab province experiences the highest number of crashes in Pakistan (Batool et al., 2018), and Rawalpindi, being the third biggest city in the province, has only one bus service i.e., bus rapid transit (BRT) operating under the management of the city's local government. BRT currently serves on a single route, particularly on Muree Road, and accommodates ridership of approximately 125,000 passengers daily. Rapid urbanization and increased vehicle use in Rawalpindi's bustling areas contribute to the intensity of RTCs (Arain et al., 2017). Due to limited public transport on feeder routes, people have no choice but to use private vehicles, ride-hailing services, and 3-MR, which are operated privately (Priye & Manoj, 2020). The auto-rickshaw known as Qingqi (pronounced "Chingchi"), is a popular type of rickshaw in Pakistan. These three-wheeled vehicles are powered by small petrol engines with a displacement of 100cc to 200cc. Three-wheeled motorized rickshaw (3-MR) play a significant role in Pakistan's transportation system, particularly in urban areas where congested streets necessitate rapid and versatile modes of transit. The 3-MR ubiquity is a defining aspect of the urban landscape in Pakistan's bustling district of Rawalpindi. Numerous commuters rely on these small, adaptable cars as their main mode of transportation because they offer last-mile connectivity and satisfy important mobility gaps in the urban environment (Navid Tahir et al., 2015). The 3-MRs are widely used in Rawalpindi, which emphasizes their socioeconomic significance and the necessity of a thorough spatial analysis to comprehend their distribution, trends, and implications for road safety. In fiscal year 2022-2023 about 17653 3-MR were sold in Pakistan, which is significantly lower than the previous year of 2021-2022 in which 32774 units of 3-MR were sold (Pakistan Automotive Manufacturers Association, 2023).

Individuals without car driving experience often turn to operate 3-MR as taxis as a means of daily financial support. The majority of these drivers are younger and middle-aged individuals hailing from regions marked by relatively lower socioeconomic status. As a result, this trend has led to increased crashes involving 3-MR, raising concerns about road traffic safety (Mirzaei et al., 2014). The structure design of 3-MR lacks basic safety elements (Srikanth & Prakash, 2009), and its open design with no doors and lack of seat belts increases the chances of injuries even with a minute crash (Hambissa et al., 2022).

The 3-MR is a popular public transportation mode in almost every Pakistani city, especially for short trips with passenger carrying capacity of four to six (Starkey et al., 2021) and most 3-MR have a moderate physical condition (Navid Tahir et al., 2015). **Figure 1.1** shows the commonly used variants of 3-MR known as ching-chi and autorickshaw (Jaiswal et al., 2024; Pervez et al., 2024). Younger Drivers and high speeds of 3-MR are more likely to result in severe injury crashes (Ijaz et al., 2021a). Drivers of 3-MR are among the highest violators of traffic rules which increases the risk of crash (Dandona et al., 2005; Jayatilleke et al., 2015).

Figure 1.1: Two commonly used variants of 3-MR: ching-chi (left) and autorickshaw (right)

Convenience and accessibility are key factors that make Rickshaws attractive to daily wage earners, working professionals and students. Rickshaws are quite essential and used as a source of transport to many individuals especially those who are of a low income. They provide a lifeline to jobs, education, health care and market for goods and services (Gadepalli et al., 2020). Despite this, rickshaws are highly preferred, though they face many challenges as well. Ever rising rate of urbanization and traffic density in the cities has given rise to concerns on road safety and pollution in Pakistan (Hissan et al., 2023). The original design and many traditional rickshaws nowadays have many unguarded features that hamper safety more, and they have poor provisions for emissions control and less safety standard, which worsens air and noise pollution. Also, due to the lack of regulation in the rickshaw sector, incidents like reckless driving are common, lack of any form of supervision means that risky driving behavior and lane changing drastically raise the crash probability. (Phun et al., 2018).

Proactively using Geographic Information System (GIS) tools enables early detection of emergency hotspots and reduces resource consumption. The GIS-based temporal analytical techniques offer local governments and policymakers a cost-effective decision-support system (Deshpande et al., 2011). Spatial analysis and statistical methods within a GIS context for identifying crash hotspots have been effective (Hazaymeh et al., 2022). Intricate connections between roadway conditions, traffic patterns, and human-made elements impact road safety, highlighting the need for integrated approaches to improve it effectively (Kumar et al., 2023). The GIS revolutionizes analytical procedures by linking attribute data to spatial data, describing an event's positional features, and visualizing data on maps to aid emergency management (Budzyński et al., 2018). Emergency management decisions are enhanced by GIS, which effectively visualizes geographically attributed data on maps (Milenković & Kekic, 2016).

Road traffic accidents continue to pose a substantial threat to public health, affecting numerous individuals with a wide range of injuries. Creating precise prediction models for the severity of injuries sustained in traffic events is essential for improving emergency response strategies and road safety legislation. Researchers have used various techniques to predict and know the cause for different types of road crashes. Forecasting the severity of crash injuries is a highly promising area of study within the field of traffic safety. Most statistical models are traditional, they have assumptions and relationships that have to be achieved before the models are implemented, this may lead to inaccurate results. Techniques such as Multinomial logit model (Hussain & Shi, 2022) regression models (Goel, 2018), random forest, logistic regression (Jacinto et al., 2023), artificial neural network, deep neural network, recurrent neural network (Amin, 2020; García de Soto et al., 2018; Rezapour et al., 2020; Sameen & Pradhan, 2017) have delivered promising results.

Machine learning algorithms have made substantial advancements in recent years, enhancing their capacity to forecast results and analyze large datasets. It has facilitated the provisions of useful information to the professionals and decision makers in the area of transport safety. Neural networks show good predictive prowess across various fields including transportation safety. Consequently, due to their capability of recognizing patterns and relations within the information given, they have great skills in anticipating rather comprehensive trends with regard to the results of traffic accidents. As for the degree of injuries severity estimation it is crucial not only in the frames of statistic calculations but in the perspective of emergency response, resource allocation and protective measures, etc. To enhance the effectiveness of the model, it is required that the selected model is capable of dealing with the complex characteristics of traffic crashes, which are influenced by multiple factors.

1.2 Problem Statement

The traffic accidents in Pakistan remain a great menace to individuals and safety of the people. Hence the nation is at a bigger risk of road safety threats associated with high crash, injury and fatality rates. 3-MR are involved in significant number of crashes in Punjab province and Rawalpindi being the third biggest city of Punjab also experience these crashes. 3-MR are preferred by public because of their rapid movement in congested areas and cheaper fares. However, Major challenges encountered yet rickshaws remain to be important transports used in Pakistan and reveals the flexibility of the transport system of the country and its cities. It is thus imperative that efforts be made for the urgent improvement of traffic safety regarding 3-MR crashes and reduced effects of these crashes on our society. Pakistan has a capability to enhance the road safety and to reduce the prevalent social/economical cost of traffic accidents through certain focused interventions and raising the public awareness. In order to improve road safety, it is crucial to address key risk factors of 3-MR crashes and implement a comprehensive strategy that includes enhancing the road network, enforcing vehicle safety regulations, ensuring strict traffic law enforcement, and providing proper post-crash care.

1.3 Scope of Research

This research analyzes the distribution, clustering patterns, and temporal dynamics of 3-MR crashes in Rawalpindi district. The study shed light on the spatial aspects of 3-MR crashes. This research also employed machine learning algorithms to study risk factors that define the crashes involving 3-MR and the correlation between the crashes, and the injury severity of crashes. The research findings underscore the need for targeted interventions focusing on crash hotspots and identifications of factors effecting crash severity ensuring effective resource allocation and evidence-based policies to enhance road safety and urban mobility giving useful outcomes for transportation planners and policymakers.

1.4 Objectives of Research

This research has been conducted to achieve the following objectives:

- Identify traffic crash hotspots of 3-MR by considering spatial aspect of crashes.
- To locate and distinguish Emerging hotspots of 3-MR crashes incorporating spatial as well as temporal aspect.
- Develop a machine learning model for injury severity prediction to accurately classify cases.
- Analyze impact of contributing factors in predicting the injury severity of 3-MR traffic crashes.

1.5 Thesis Organization

Chapter 1 presents the introduction about the topic "Traffic Crash Analysis of Three-Wheelers in Rawalpindi: Prediction Modeling using Machine Learning and Emerging Hotspot Analysis". Details about road traffic crashes, urban public transport and role of 3-MR, use of GIS in traffic safety, and machine learning approaches in predicting crash injury severity are explained in this section.

Chapter 2 consists of literature review, in which significant studies from past about road safety in developing countries, 3-MR road safety, use of hotspot analysis approach for traffic crashes, use of different machine learning approaches in traffic safety, and factors contributing to injury severity are discussed in detail.

Chapter 3 discuss about the methodology that is adopted in order to determine hotspots related to traffic crashes, and machine learning approach to predict injury severity and contributing factors.

Chapter 4 comprises of detailed analysis and results of hotspots analysis approach as well as machine learning prediction modeling for injury severity. Results give detailed insights into spatio-temporal patterns of traffic crashes and factors contributing to injury severity.

Chapter 5 provides detailed conclusion withdrawn from analysis and results about the study area and presents insights that are helpful in guiding authorities for targeted road safety interventions.

CHAPTER 2: LITERATURE REVIEW

2.1 Road Safety in Developing Countries

A region's socioeconomic status affects crash rates (Constant et al., 2008), which are significantly different between developing and developed nations. Despite having a total of 54% of global vehicles, developing countries suffer 90% of global traffic accidents (World Health Organization, 2018). With economic growth, countries improve their public health and safety, which is lacking in developing countries, and road safety is unaccounted for due to limited resources (Bishai et al., 2006). Lack of awareness of road safety and perception of law-abiding is also not regarded by road users in developing nations, which results in severe traffic crashes (Jadaan et al., 2018). A study on drivers in a developing country revealed risky driving behavior increases the risk of crash by 50%. Authors also highlighted that knowledge and education about safe driving had no effect on reduced crashes unless drivers had safer driving attitudes (Mirzaei et al., 2014). Another research suggested stricter law enforcement can lead to better attitudes of drivers towards safety. Findings also showed that driving experience and availability of safety measures were likely to reduce the crash risk (Chakraborty & Maitra, 2024). A study conducted on traffic crashes in Pakistan highlighted causes of the rising RTC rate include inadequate safety measures, speeding, cell phone use, driver recklessness, limited traffic regulation knowledge, and ineffective enforcement by traffic authorities (Imran & Nasir, 2015). (Deshpande et al., 2011) also pointed out that the fatality rate from traffic crashes has increased in developing nations due to inadequate road infrastructure, poor vehicle conditions, and insufficient enforcement. High-income nations have seen a decline in traffic crash deaths and injuries over the past thirty years, a trend not observed in low- and middle-income countries.

2.2 Road Safety of Three Wheeled Motorized Rickshaw

Victims of traffic accidents, including pedestrians, cyclists, and motorcyclists, account for almost 50% of all casualties (World Health Organization, 2018). 3-MR were involved in a significant number of road crashes in Punjab province between 2011-2013 (Navid Tahir et al., 2015). The 3-MRs are used as low-cost taxi services due to the cheaper rates in developing nations like Pakistan, India, and Bangladesh. These can accommodate four to six persons depending on the type of rickshaw, but these vehicles are negatively perceived among the majority of road users due to driver's behavior and contribution to road congestion. 3-MRs are not widely used in rural areas but are an important part of the traffic in urban environments of developing countries (Starkey et al., 2019).

The actual number of crashes and injuries are significantly higher than reported figures (Subhan et al., 2021). In past not enough studies have been done on safety of 3-MR, which leaves a gap in research. Few researchers have done research on 3-MR (Ijaz et al., 2021b; Navid Tahir et al., 2015; Pervez et al., 2024; Starkey et al., 2021). (Ijaz et al., 2021b) used machine learning classifiers on crash data of 3-MR to explore factors contributing to injury severity. Their research suggested factors like weather, time of day, and driver characteristic have significant impact on 3-MR crash injury. Using statistical approach. (Pervez et al., 2024) also found age, time, weather, type of crash as important variables in determining 3-MR crash injury severity. The primary cause of injury for the majority of patients, accounting for nearly 56% of cases, was a collision with a moving vehicle. This was followed by falls from rickshaws (Meena et al., 2014). 3-MR small size and maneuverability allow them to stop anywhere on the road, impeding traffic flows. Lack of safety protection leads to poor outcomes for passengers involved in collisions (Starkey et al., 2019). The high rate of fatalities involving 3-MR has raised questions about their structural integrity during crashes. (Jurangpathy & Rasika, 2022) examined 3-MR crashworthiness and suggested design changes to improve passenger safety. Findings of research carried out by (Navid Tahir et al., 2015) are crucial for manufacturers and policymakers to consider when formulating safety standards for 3-MR. In Pakistan regarding vehicle characteristics, most 3-MR are in moderate or poor physical condition, with only a small percentage in good condition. Crashes involving rickshaws were determined to have a higher probability of resulting in non-injury crashes (Ullah et al., 2021). This conclusion is easily understandable because the rickshaw has comparatively lower speed and weight attributes. But research conducted by (Batool et al., 2018) indicated that women have a greater risk of deaths compared to male drivers. Rickshaws and cars are more prone to deadly accidents compared to other types of vehicles due to their significant presence on the roads.

Regulation is also a key issue for 3-MR. They are often not considered authorized vehicles, banned from some routes and locations, and operated by deprived, under-age, risk-taking youths with little training (Starkey et al., 2019). To increase the road safety in Pakistan measures such as dedicated bike lanes, improvement of road infrastructure, enforcement of traffic laws, and raising public awareness are needed (Mansoor et al., 2023). Strict enforcement of seatbelt laws, lowering of speed limits, and the introduction of detailed road safety audits are recommended to mitigate road crash injuries (Subhan et al., 2021).

Authors also highlighted the difficulties in accurately collecting road crash data in Pakistan, specifically the presence of inconsistencies and inadequate reporting in current systems. The lack of road safety policy and weak law enforcement in Pakistan, contribute to low compliance with traffic laws (Navid Tahir et al., 2015). It emphasizes the need for authorities to manage 3-MR related transport and road safety issues. Limited amount of literature suggests to further investigate the safety of 3-MR.

To understand the nature of traffic crashes for 3-MRs, a study in India found front collisions of 3-MR to be responsible for severe injuries, and front seat passenger also has a higher chance of suffering fatal injury (Schmucker et al., 2011). Lack of stability, basic safety elements, crashworthiness, and unsafe structure also impact crash injury outcomes (Sindha et al., 2015; Srikanth & Prakash, 2009). A study investigated crash characteristics of 3-MR in Sri Lanka and found crashes to be dominant during daytime, weekdays, and while driving on rural highways (Amarasingha, 2015). (Nishantha Vadysinghe et al., 2018) carried out a study to explore injury patterns of 3-MR crashes and found most of the victims were males, the majority of fatalities occurred as a result of injury to the head, and overturning of 3-MR was the dominant cause of the crash.

2.3 Hotspot Analysis Approach in Traffic Crash Assessment

A hotspot in the context of traffic crashes refers to a place where the dominant clustering of traffic crashes occurs. Due to land use and socio-economic factors, some places may experience a higher number of crashes than others. A study employed clustering algorithms to examine the spatiotemporal patterns of traffic crashes, using the Global Moran I index and Getis-Ord Gi* analysis to identify significant crash clusters along road segments

(Hazaymeh et al., 2022). (Butt et al., 2017) identified hotspots and high-risk regions for RTCs in Rawalpindi. Their analysis included spatial autocorrelation (Moran's I test), standard deviational ellipse (SDE), and hotspot (Getis-Ord Gi*) analysis, revealing crash clusters near highways, airports, schools, and hospitals. Their research highlights the importance of incorporating GIS in emergency management to improve emergency response. In Abu Dhabi, spatial autocorrelation analysis identified and ranked high-density crash sites in the central business district, showing the proficiency of the Getis-Ord Gi* approach (Alkaabi, 2023). Methods like Kernel Density Estimation (KDE), Network Kernel Density Estimation (NKDE), and ordinary kriging techniques are used to identify crash hotspot sections, whereas NKDE can be most effective for smaller segments (Bisht & Tiwari, 2023). (Çela et al., 2013) employed the NKDE approach pinpointed cluster locations, while the Network K-function explored their presence. In South Korea, a spatiotemporal hotspot analysis revealed that fatal two-wheeler crashes were concentrated in the densely populated capital, with speeding and running red lights as major contributing factors (Tamakloe, 2023). A study on the spatiotemporal distribution of incidents involving pedestrians, public transit, and unconventional transportation found collision hotspots in central business districts, airports, and ferry ports. Research suggests different temporal tendencies between affluent and developing nations, using planar kernel density analysis to evaluate crashes involving various transportation modes in terms of space and time (Saha et al., 2021).

Emerging hotspot analysis helps to detect spatiotemporal patterns of traffic crashes. A space-time cube, which contains information about the location and time of the crash in three dimensions, is used to perform emerging hotspot analysis. A study combined space-time cube analysis, spatial autocorrelation analysis, and emerging hotspot analysis to investigate traffic crash evolution and pinpoint hotspots, identifying areas with prevalent crash types (Cheng et al., 2019). In North Dakota, using data of single-vehicle lane departure crashes, (Khan et al., 2023) used spatiotemporal techniques like NKDE and emerging hotspot analysis using a space-time cube to identify crash hotspots and found that 90% of fatal crashes occurred on rural highways and several counties also identified as crash hotspots. The NKDE results indicated that most of the crash clusters were located on junctions, curves, and intersections. Research on articulated heavy truck incidents in Western Australia used spatiotemporal methods to detect emerging hotspots, revealing most crashes occurred near the Perth metropolitan region (Gudes et al., 2017). A study on single-vehicle crash patterns in Western Australia between 1999 and 2008 found significant spatial and temporal disparities, particularly in urban centers and along major roads (Plug et al., 2011). (Mohammed et al., 2023) explored spatiotemporal road traffic collisions in Qatar, finding that crashes were mainly concentrated in the central-eastern region and were linked to driving behavior. The research indicated that weekday crashes in 2019 were more clustered than in 2015. Researchers have suggested a comprehensive examination of RTCs using GIS and time series analysis to reveal patterns and emphasize the influence of seasonal and local variables on traffic crash rates (Mahmood et al., 2022). Accurate identification of collision risk hotspots enables urban transportation authorities to implement effective safety measures (Wu et al., 2023). Digital technologies like artificial intelligence (AI), machine learning, and GIS can enhance road safety by automating processes and reducing human bias (Eskandari Torbaghan et al., 2022).

2.4 Machine Learning Approaches to Predict Injury Severity of Traffic Crashes

For RTCs researchers have used many methods to predict injury severity of crashes. (García de Soto et al., 2018) used data from 2009 to 2012 of national roads of Switzerland and found ANN to be feasible method for predicting traffic crashes. (Rezaie Moghaddam et al., 2011) used series of ANNs to model the factors that are related to traffic crash. (Amin, 2020) employed back propagation artificial neural network (ANN) to simulate the variables influencing traffic accidents involving older drivers. Study concluded that the most crucial factors contributing to accidents among older drivers were the purpose of the journey, time of day, lighting conditions, pedestrian crossings with physical interventions, intricate roadway geometry and severe weather conditions. (Alkheder et al., 2017) used data from 2008 – 2013 traffic crashes to predict injury severity using ANN, and achieved an accuracy of 75% in their prediction model, furthermore they used "k means algorithm" to divide data into three clusters which also improved the prediction accuracy of model. To compare the results, they used ordered probit model which gave accuracy of up to 60% which was less than ANN. A comparative study evaluates three distinct machine learning techniques for forecasting the severity of injuries. Results indicate that the Decision Jungle algorithm outperforms other classification algorithms like Random Forest and Decision Tree (Bokaba et al., 2022). (Singh & Suman, 2012) found that heavy vehicles are involved in approximately 48% of accidents, followed by two-wheelers, cars, and buses. (La Torre et al., 2019) used approach based on Highway capacity manual to develop a crash prediction model. Highway crash data from 2009 to 2015 was used to predict injury severity of traffic crashes. Recurrent neural network obtained an accuracy of 71%, Multilayer perceptron and Bayesian logistic regression models had lower accuracy of

65.48% and 58.30% respectively (Sameen & Pradhan, 2017). A study examined 2,430 motorcycle crash records collected from a mountainous region in the United States throughout a span of 10 years using a comparative analysis of various deep neural networks (DNNs), such as deep belief network, recurrent neural network (RNN), multilayer neural network, and single-layer neural network. The analysis involves the utilization of feature reduction techniques and evaluation metrics like the area under the curve and confusion matrix to assess the performance of these models. But the models couldn't achieve satisfactory accuracy for underrepresented class i.e. Severe/Fatal and only favored overrepresented class (Rezapour et al., 2020). Some crashes also occur as a result of previous accident. (Wang et al., 2019) used back propagation neural network to forecast the time interval between such accidents. Extensive literature review shows artificial neural networks have proven to be promising tool in case of predicting crash injury severity.

2.5 Contributing Factors to Injury Severity from Past Studies

These developed models then identify factors that contribute to crash or injury severity. A study used regression model that discovered walking, cycling, and 3-MR were linked to a reduced possibility of road fatalities in Indian states, while cars, motorcycles, and buses are associated with higher probability (Goel, 2018). A study carried out to predict the injury severity, recommended gender of the driver, type of the vehicle, use of the seat belt, location of crash, and the type of area as important factors (Abdel-Aty, 2003). To find the relation between temperature, rainfall, health worker density index, and road traffic deaths, a study used RTC data of Pakistan from 1985 to 2016. It found that a 1% increase in temperature led to a 3.6% increase in road fatalities (Ali et al., 2020). Frequency of RTCs was found positively related with weather parameters such as rainfall, extreme cold, fog, and heat (Hammad et al., 2019).

A study developed a statistical model that estimates the death rates in RTCs with reference to socio-economic factors and identified that the density of roads, maximum allowable speed, the observance of law on child restraint in motor vehicles and existence of a funded lead agency, are some of the factors that influence the crash fatality rate (Ahmed et al., 2016). (Ospina-Mateus et al., 2021) have identified time of crash, weather, road state, and the number of victims as some of the variables of accidents. They also recommended ways to enhance road safety, a critical issue that keeps on rising in the society. Underreporting of RTCs is also a concern, (Rolison et al., 2018) stressed the importance of reviewing and updating accident report forms to include all contributing causes to road accidents. It is also possible for accident reporting practices to vary depending on the characteristics of the driver, such as the driver's age and gender (Pourroostaei Ardakani et al., 2023). A study emphasized the significant risk factors, which include the age of the driver, the month and day of the week, and the type of vehicle. Additionally, there is a correlation between young drivers, weekends, and the summer season with increased accident severity (Khattak et al., 2022). Age group of 21-40 years suffered most from the 3-MR related crashes (Awan et al., 2020). Drivers below the age of 25 years are at a higher chance of gaining minor injury and at a lower risk for other injuries. Young drivers are inexperienced and not cautious enough hence the high possibility that they will engage in an accident. However, elderly drivers tend to have compromised physical and mental health than younger drivers thus young drivers have an appreciable ability to identify and manage risks than the older drivers (Kamti & Iqbal, 2021; Ullah et al., 2021).

A sizable fraction of the drivers had driven while in possession of an invalid driver's license, engineering countermeasures and education are probably more successful than governmental regulations and enforcement actions (Somasundaraswaran & Tay, 2006). It is recommended to implement more efficient safety management measures, such as variable speed limits and ramp metering, that are customized for different types of road segments and traffic flow conditions on expressways in order to decrease the number of accidents (Kwak & Kho, 2016). Factors such as distracted driving, collisions with pedestrians or trucks, and female riders were also found significant contributors to crash severity (Mansoor et al., 2023). (Hussain & Shi, 2022) Conducted a study to analyze the impact of several predictor variables on road crash data from 2013-2017 among functionally classed automobiles in Pakistan, using a multinomial logit model that identified several key factors which contribute to RTCs, including reckless driving, driver fatigue, excessive speed, tire blowouts, and bad road conditions. These factors had different levels of impact on motorcycles, non-commercial cars, and commercial vehicles.

2.6 Literature Review Summary

In the context of 3-MR crashes, extensive literature review suggests no work has been done regarding pointing out the locations where these 3-MR crashes are prevalent. Further some research has been done regarding statistical modeling of 3-MR crashes but no research has considered effect of emergency response time on 3-MR crash injury severity. This study fills this gap by conducting a detailed crash hotspot analysis of 3-MR and implementing techniques like hotspot analysis using Geti Ord Gi* statistics and emerging hotspot analysis using the space-time cube technique. Geti Ord Gi* statistics use spatial aspect to detect hotspot patterns while emerging hotspot analysis differentiates between different crash patterns while using spatial as well as temporal aspects of 3-MR crashes, and categorize crash hotspots into different categories, which helps better to understand the frequency and occurrence of such incidents. Further key variables like emergency response time and their impact on injury severity of these crashes is analyzed using artificial neural networks, which gives key insights regarding 3-MR crashes.

CHAPTER 3: METHODOLOGY

3.1 Study Area and Data

Crash data was acquired from Rescue 1122 Rawalpindi District, an emergency response service established by the Punjab Emergency Service Act 2006. Crash data for a period of one year and four months from January 2022 to April 2023 was collected. Data consisted of all types of vehicle crashes, from which only cases of 3-MR crashes were filtered out. During this time period 1213, 3-MR crashes were reported .1420 casualties occurred as a result of these crashes out of which 298 sustained severe injuries while 1198 had minor injuries. The dataset included 3-MR crash locations, crash ID, Date and Time of Crash, demographic variables like age, gender, educational qualification of victims and injury severity, cause of accident, victim type (driver, passenger or pedestrian), emergency response time, and number of vehicles involved. Meteorological data was collected from Pakistan Meteorological Department, which included weather conditions (i.e. cloudy, sunny, or rainy) and Daily temperature as shown in **Table 3.1**. Crash locations were noted in the data as addresses to the nearest landmarks and not in coordinates. To obtain crash location coordinates, location addresses were geocoded through geocoding APIs, after which, for each crash, location coordinates were again put in Google Earth to place those coordinates on road networks that were not. Crash data was imported into ArcGIS software from an Excel CSV file for analysis, revealing crash locations throughout the district when plotted on a map. To facilitate analysis, coordinates were converted into a projected coordinate system, transforming them from World Geodetic System (WGS)-1984 to (UTM) Universal Transverse Mercator (linear metric units) system. The Shapefile of the road network was obtained through Open Street Maps for the study area. **Figure 3.1** shows the 3-MR crash locations along with the road network of Rawalpindi for the study area. Software like MS Excel, Google Sheets and Geocoding APIs were used to handle the data, while ArcGIS Pro was used to perform spatial and temporal analysis of crash data. SPSS and python were used to perform prediction modeling.

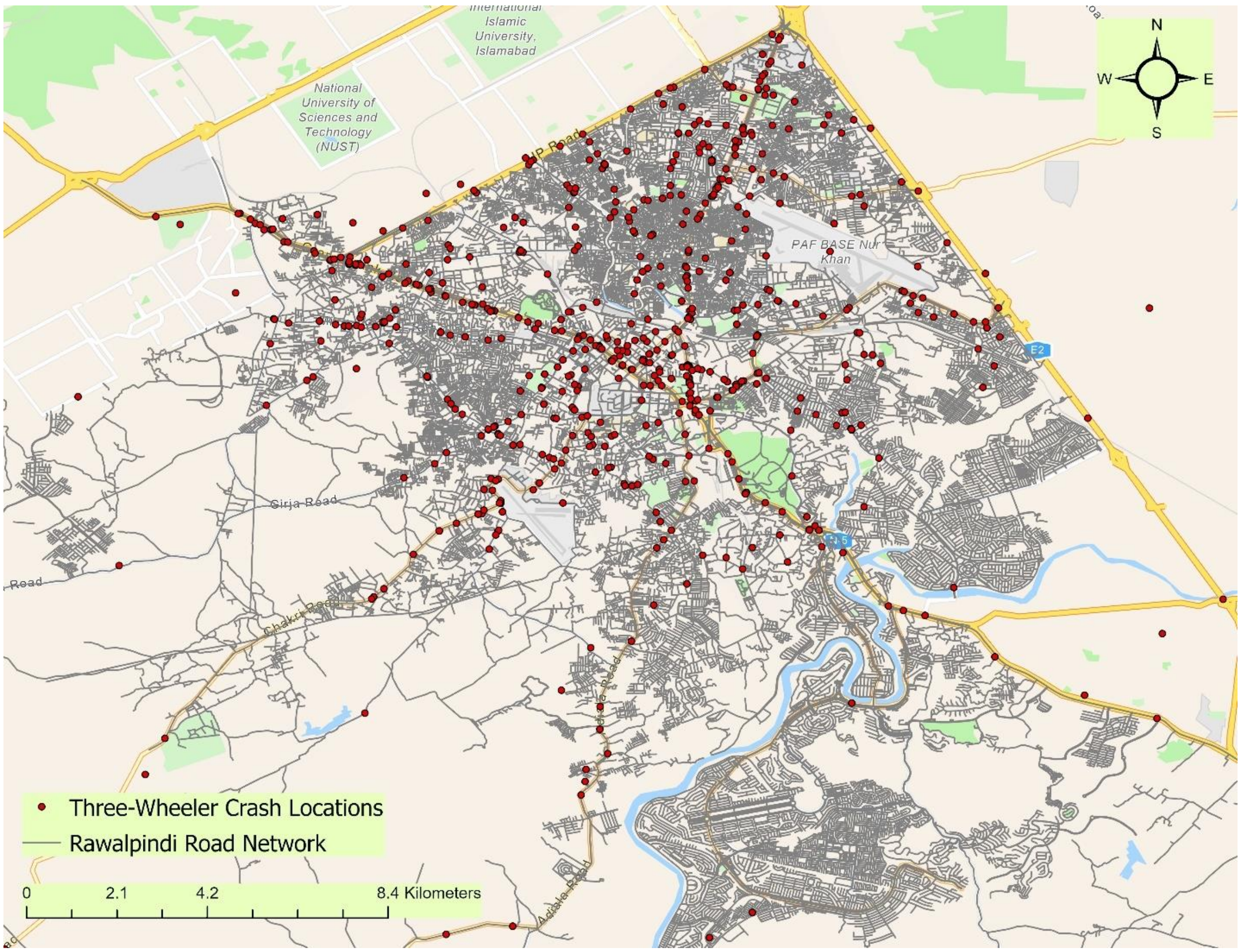

Figure 3.1: Crash Locations on Rawalpindi Road Network

Table 3.1 Data Variables

Factors	Category
Response Time	Continuous
Average Temperature	Continuous
Number of victims	Continuous
Day	1=Weekend, 2=Weekday
Season	1=Spring, 2=Summer, 3=Autumn, 4=Winter
Time	1=6am-12pm, 2=12pm-6pm, 3=6pm-12am, 4=12am-6am
Peak Hours	1=Morning Peak, 2=Evening Peak, 3=Off-Peak
Gender	1=Male, 2=Female
Age	1=<18, 2=18-30, 3=30-45, 4=45-60, 5=>60
Victim Type	1=Pedestrian, 2=Passenger, 3=Driver
Education	1=No Education, 2=Basic Education, 3=Matric, 4=Inter, 5=Graduate
RTA Cause	1=Over speeding, 2=Careless driving, 3=Wrong Turn, 4=One Wheeling, 6=Tire Burst, 7=Others
Vehicles Involved	1=Single Vehicle, 2=2 Vehicles, 3=More than 2 Vehicles
Weather	1=Sunny, 2=Cloudy, 3=Rainy
Month of year	1=January, 2=February, 3=March …. 12=December
Time in Meridian	1=AM, 2=PM
Light Conditions	1=Day, 2=Night
Day of week	1=Monday, 2=Tuesday, …. 7=Sunday

Figure 3.2: Research Flow Chart

3.2 Global Spatial Autocorrelation

Moran's *I* is a correlation coefficient that quantifies the geographical autocorrelation of data, assessing how closely related an object is to those around it. If two items are drawn to or away from each other, it indicates that the observations are not independent, violating a fundamental tenet of statistics. Testing for autocorrelation is crucial because its presence invalidates most statistical tests.

Global Moran's *I* analyze the pattern of crash locations and test the null hypothesis that "attributes are randomly distributed across the study area". Rejection of this hypothesis indicates spatial autocorrelation (Ord & Getis, 1995).

Mathematically, Moran's Index can be represented as:

$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1\ j\neq i}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{S_0 \sum_{i=1}^{n}(x_i - \bar{x})^2} \ \forall_i = 1, \ldots, n \land \forall_j = 1, \ldots, n \qquad (1)$$

Where,

w_{ij} represents the spatial weight indicating the Relation between neighboring features i and j.

x_i is the location at which the feature value exists,

x_j is the location of a feature at an adjacent point,

S_0 represents the total of all spatial weights,

$\bar{x}$ represents the average value of the characteristics, whereas N represents the total count of places.

Moran's Index ranges from -1 to 1. A value greater than 0 implies a positive correlation in geographical distribution, indicating clustering of similar values. A value near 0 suggests a random spatial distribution. A negative value indicates significant variation within the same area, known as dispersion (an ideal grouping of varying values), as shown in **Figure 3.3**. The calculated values are then transformed into a z-score. A positive z-score indicates neighboring features have similar values, while a negative z-score shows adjacent features have different values (Cheng et al., 2019). If the p-value is statistically significant, it allows for rejecting the null hypothesis.

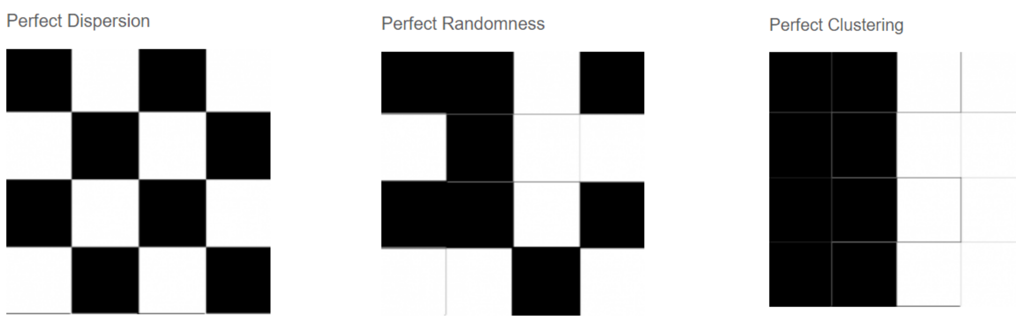

Figure 3.3: Distribution of points in space (Glen, 2023)

3.3. Incremental Spatial Autocorrelation

Instead of selecting a single distance threshold for spatial autocorrelation analysis, Incremental Spatial Autocorrelation (ISA) allows you to systematically investigate multiple distances, helping to choose the most effective threshold for a specific dataset. This results in more precise and meaningful outcomes.

ISA produces a curve or chart displaying autocorrelation values at different distance intervals, quickly highlighting significant geographical patterns or clusters in the data. The distance used greatly impacts how spatial autocorrelation is shown (Khan et al., 2023). The

Incremental Spatial Analysis tool in ArcGIS Pro can be used to find the optimum distance for the best clustering of traffic crashes. This tool generates Z values for increasing distances, with higher Z scores indicating prominent clustering, as shown in the chart peaks **Figure 3.4**.

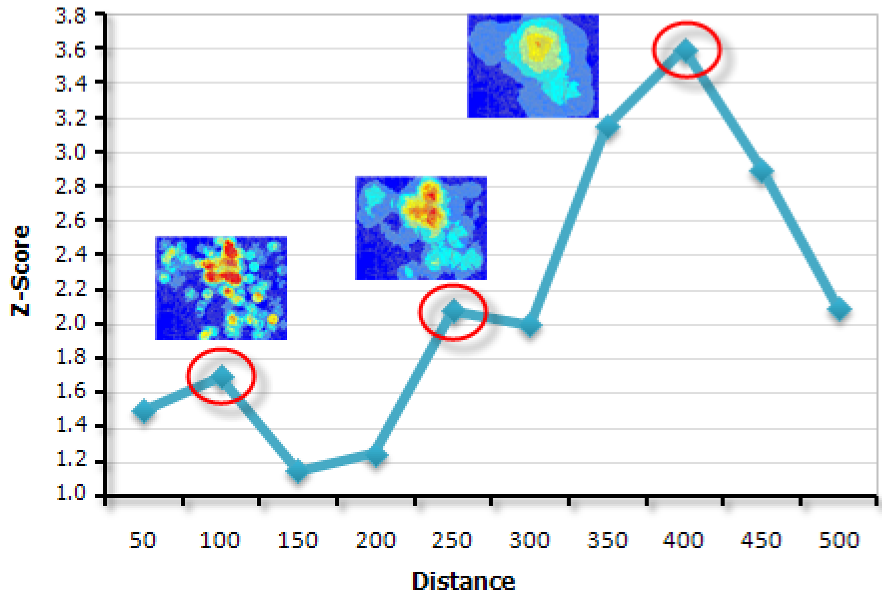

Figure 3.4: Spatial Autocorrelation by Distance (ESRI, 2023c)

In the case of multiple peaks, using the first peak value as the distance band is recommended for further analysis. To determine the initial distance for ISA, the "Calculate Distance Band from Neighbor Count" tool in ArcGIS Pro helps determine the average distance at which each location has at least one neighboring point (Khan et al., 2023).

3.4. Local Spatial Autocorrelation

Local patterns may sometimes be anomalies that global indicators could overlook, or they could contradict the overall geographical trend (Anselin, 1995). Local spatial autocorrelation is introduced to identify significant local spatial clustering; the Getis-Ord

Gi and Gi* statistics are beneficial when spatial data are available for crash location analysis (Deshpande et al., 2011; Getis & Ord, 1992).

Hotspot analysis proves essential in pinpointing suitable interventions for areas exhibiting clustering patterns, such as traffic crashes, and comprehending the underlying causes of these clusters. It also aids in visualizing the geographical locations and extent of these clusters. The algorithm computes the Gi* statistic for each dataset feature, identifying spatial clusters where high or low z-scores and p-values are concentrated. It evaluates each feature by comparing it with neighboring features, marking it as a statistically significant hotspot if it exhibits a high value and is surrounded by other features with high values (Khan et al., 2023).

The Gi* statistic acts as a z-score, where a higher value indicates a more concentrated presence of high values, and a lower value indicates a stronger clustering of low values. Local statistics like Gi and Gi* play a crucial role in uncovering localized spatial patterns that might remain unnoticed when relying solely on global statistics (Ord and Getis, 1995).

Mathematically,

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i.j} x_j - \bar{X} \sum_{j=1}^{n} w_{ij}}{s\sqrt{\frac{[n \sum_{j=1}^{n} w_{i.j}^2 (\sum_{j=1}^{n} w_{i.j})^2}{n-1}}} \quad (2)$$

where,

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - (\bar{X})^2} \quad (3)$$

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n} \tag{4}$$

The equation represents the relationship between the traffic crash rate value (n), the property value for the j^{th} element (x_j), the spatial weight between zone i and zone j (w_{ij}), and the mean of the variable ($\bar{X}$). This weight matrix is calculated by taking the reciprocal of the distance between sites i and j. (i.e., $1/ d_{ij}$) (Mohammed et al., 2023b).

3.5. Space-Time Cube and Emerging Hotspot Analysis

Pattern analysis often utilizes optimized hot spot analysis and emerging hot spot analysis, where hot and cold zones are defined by underlying spatial associations rather than random processes, as described by Getis and Ord. Emerging hot spots enrich the dataset by incorporating a temporal dimension.

Space-time cube analysis, a 3D visualization technique, transforms spatiotemporal data into a cube format to identify spatiotemporal patterns. Crash frequency statistics are converted into Network Common Data Form (netCDF) data using an ArcGIS Pro tool that aggregates points to construct the space-time cube. The space-time cube employs the z-axis for time and the x and y-axes for geographical dimensions (**Figure 3.5a**). Bin time series are generated through spatial placement, with each vertical column representing a unit cube and containing the count of geographic occurrences within each time step (**Figure 3.5b**). These bin time series visually illustrate changes in geographic occurrences over time (Cheng et al., 2019).

After identifying emerging hot spots, each bin is assigned a z-score, p-value, and hotspot categorization. The Mann-Kendall trend test evaluates hotspot and cold spot patterns using

Mann-Kendall statistics (Mann, 1945). Throughout the hot spot analysis process, parameters affecting a space-time bin's statistical significance relative to its neighbors can be adjusted, including neighborhood distance, neighborhood time step, and spatial interconnection conceptualization.

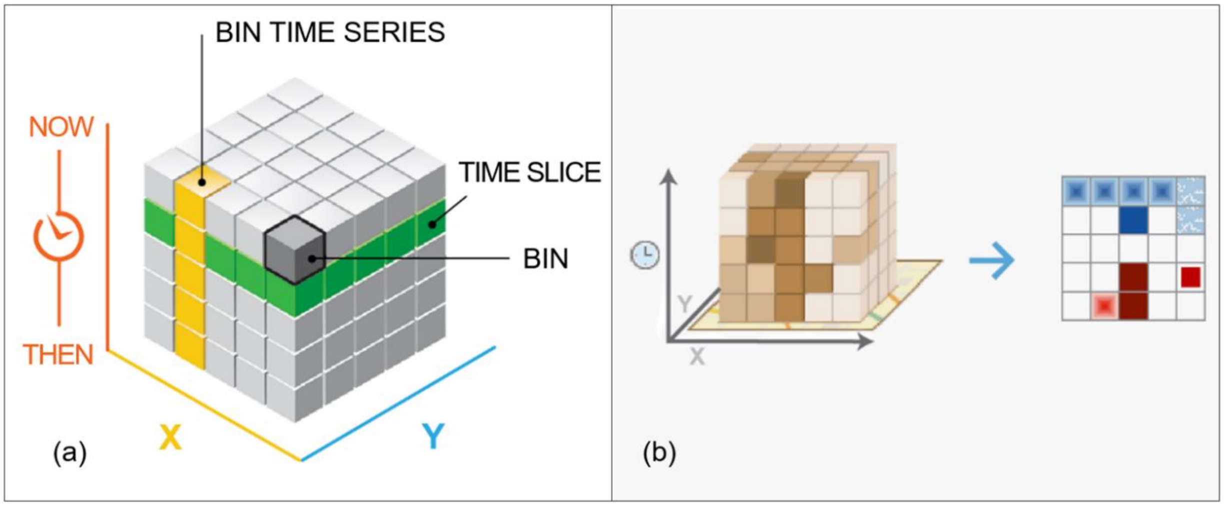

Figure 3.5: Space-time cube structure (a) Space-time bins in 3D; (b) Generated bins in 2D for emerging hot spot analysis (ESRI, 2023a)

Three key factors, neighborhood distance, neighborhood time step, and conceptualization of spatial relationships, must be controlled to analyze links between space-time bins. Neighborhood distance determines the maximum spatial distance between two bins considered as spatial neighbors, while the neighborhood time step sets the maximum time interval between two bins to be considered temporal neighbors. The conceptualization of spatial relationships is crucial and varies based on individual preferences (Harris et al., 2017).

Through emerging hotspot analysis, the results of z-score, p-value hotspots, and bin trends are categorized into seventeen groups. **Source**: (ESRI, 2023b)

Table 3.2 outlines these hotspot categories briefly. Source: (ESRI, 2023b)

Table 3.2 Types of space-time cube patterns with names and definitions.

Symbol	Name	Description
	No Pattern Detected	A location where none of the hot/cold spot patterns listed below are applicable.
	New Hot/Cold Spot	A location that is a statistically significant hot/cold spot for the final time step and has never been a previously statistically significant hot/cold spot.
	Persistent Hot/Cold Spot	A location that has been a statistically significant hot/cold spot for 90% of the time-step intervals with no discernible trend indicating an increase or decrease in the intensity of clustering over time.
	Diminishing Hot/Cold Spot	A location that has been a statistically significant hot/cold spot for 90% of the time-step intervals, including the final time step. Additionally, the intensity of clustering in each time step is decreasing overall and that decrease is statistically significant.
	Sporadic Hot/Cold Spot	A location that is an on-again off-again hot/cold spot. Fewer than 90% of the time-step intervals have been statistically significant hot/cold spots and none of the time-step intervals have been statistically significant cold/hot spots.
	Consecutive Hot/Cold Spot	A location with a single uninterrupted run of statistically significant hot/cold spot bins in the final time-step intervals. The location has never been a statistically significant hot/cold spot before the final hot/cold spot run and fewer than 90% of all bins are statistically significant hot/cold spots.
	Intensifying Hot/Cold Spot	A location in which the intensity of clustering of high counts in each time step is increasing overall and for which the increase is statistically significant (90%).
	Oscillating Hot/Cold Spot	A statistically significant hot/cold spot for the final time-step interval that has a history of being a statistically significant cold/hot spot during a prior time step. Fewer than 90% percent of the time-step intervals have been statistically significant hot/cold spots.
	Historical Hot/Cold Spot	The most recent time period is not hot/cold, but at least 90% of the time-step intervals have been statistically significant hot/cold spots.

3.6. Artificial Neural Network

3.6.1 Model Parameters

Artificial Neural Network (ANN) behaves like neurons in human nervous system, and learns the pattern accordingly. ANN was used to develop a prediction model for traffic crashes of 3-MR in Rawalpindi district, in which injury severity was considered as dependent variable and gender, time, day, age, education, peak hours, number of victims involved, victim's type (passenger, pedestrian or drive), number of vehicles involved, cause of crash, season, temperature, average daily temperature, weather conditions and emergency response time were considered as independent variables.

3.6.2 Model Structure

For this study feed-forward neural network was implemented using backpropagation training algorithm. Data was split into 70% for training dataset and 20% for testing data set and 10% for validation. Since most of the crashes resulted in Minor injuries and with second highest being multiple to severe injuries and only 0.28% being fatal. So fatal, multiple and spinal/head injuries were combined in a single class. Minor injuries comprised of 79% while Severe injuries comprised of 21%. Input layer consisted of 14 neurons which were input/dependent variables. Number of layers and neurons is trial error procedure, after which one hidden layer was chosen with 8 number of neurons. Input variables were standardized to ensure equal weights initially. Standardization improves the performance, stability, and interpretability of statistical and machine learning models. (Ma et al., 2021). Activation function hyperbolic tangent (tanh) was used for hidden layers, tanh introduces non linearity into the system and helps to catch complex patterns and produces output in range (-1,1). This helps hidden neurons to avoid saturation and enables the network to learn complex non-linear relationships in the data. Sigmoid activation function was selected for output layer. This function converts the raw output scores of the network into probabilities, with each output representing the probability of belonging to a particular class.

3.6.3 Synaptic Weights

The parameter estimates for the hidden layer correspond to the weighted connections linking the input layer predictors with the neurons in the hidden layer denoted as H(1:1), H(1:2), H(1:3), and so on. Each value in the Hidden Layer section represents the weight associated with a specific predictor's connection to a neuron in the hidden layer. The weights regulate the influence of each prediction on the activation of the corresponding neuron. The weight's size signifies the intensity of the connection, while its sign denotes the direction of this connection. Positive weights signify that greater values of the predictor enhance the activity of the neuron, whilst negative weights imply a detrimental effect. Moreover, the inclusion of the bias term in the hidden layer adds a constant value that enhances the overall activation of the neurons, hence impacting the final output. This allows the model to effectively capture intricate patterns and non-linear connections in the input data, which is crucial for making precise predictions about the severity of injuries in traffic accidents.

3.6.4 Synthetic Minority Oversampling Technique

Since Fatal or severe accidents of 3-MR occur rarely, so in our case this class was underrepresented and most of the crashes were minor accidents which lead to over fitting in our model. Imbalanced datasets can lead to biased models that favor the majority class and perform poorly on minority class instances (Bader-El-Den et al., 2019). Similar class imbalance was encountered by (Rezapour et al., 2020) which resulted in model favoring only overrepresented class for which they suggested to use Synthetic Minority Oversampling Technique (SMOTE). Instead of creating duplicate cases for the underrepresented class, this technique creates minority incidents at random intervals from the existing dataset.

The SMOTE algorithm works by creating synthetic samples for the minority class based on the feature space similarity of existing minority class instances. For each minority class instance, SMOTE selects one or more of its nearest neighbors (usually k-nearest neighbors) and generates synthetic samples along the line segments connecting those neighbors (Umam Syaliman, 2021). SMOTE was implemented in python using libraries such as imbalanced-learn.

3.6.5 Performance Metrics

Evaluation of the ANN model for predicting the level of traffic crash injuries incorporated performance measures, including the confusion matrix, precision, accuracy, F1 score, sensitivity (recall), specificity, and the Receiver Operating Characteristic (ROC) curve. The confusion matrix results show the performance of the model in the training, testing, and validation datasets, given useful information on the capability of the model to estimate the severity of traffic injuries accurately. This also help identify key results that allow for assessment of the model’s effectiveness and suggestions for the possible enhancements in terms of expected accuracy. Precision is a metric of the proportion of true positive (TP) predictions among all of the model's positive predictions while accuracy is the percentage of all successfully predicted positive and negative cases among all the predictions.

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

The F1-score is the harmonic mean of precision and recall. Sensitivity (recall) also known as true positive rate (TPR) is calculated as true positives (TP) over sum of true positives (TP) and false negatives (FN). If values of sensitivity are high this means model is performing well. Specificity also known as false positive rate (FPR) is calculated as false positives (FP) over sum of false positives (FP) and true negatives (TN). ROC curve illustrates trade-offs between true positive rate (TPR) and false positive rate (FPR). TPR (also known as Sensitivity) is plotted on y-axis and FPR which is 1-Specificity on x-axis. After plotting curve, Area under the curve (AUC) is calculated. Diagonal straight line which has AUC value of 0.5 depicts model is making random predictions while the curve deviating from diagonal line towards top left with AUC values between 0.5-1, indicates model's ability to accurately predict positive cases without falsely classifying negative cases as positives.

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$Specificity = \frac{TN}{TN + FP} \tag{8}$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{9}$$

$$True\ Positive\ Rate = \frac{TP}{TP + FN} \tag{10}$$

$$False\ Positive\ Rate = \frac{FP}{TN + FP} \tag{11}$$

3.7 Chapter Summary

This chapter discussed in detail about the methods that are adopted in order to achieve the objective of the research. As this research consisted of two parts i.e. Hotspots analysis using spatial and temporal data of 3-MR crashes while second part being prediction modeling for injury severity of traffic crashes. For hotspot analysis ArcGIS is used in which methods like Global and Local spatial autocorrelation, Incremental spatial autocorrelation, and Emerging hotspot analysis using space time cube is done. For prediction modeling ANN with back propagation is used. Performance metrics are analyzed is detail in order to check reliability of model. Further variables are checked for importance in prediction injury severity using sensitivity index. Results of these methods are discussed in detail in next chapter.

CHAPTER 4: RESULTS AND DISCUSSIONS

4.1 Introduction

This section presents the outcomes of the methods and techniques applied in this study. Results of each analysis is presented under heading of that method. Results of crash hotspots showed significant clustering of crashes and pinpointed the locations. These identified locations help to understand root causes of such crashes according to their social and economic use. Machine learning model using artificial neural network is also analyzed using various performance metrics in this chapter. Performance metrics like accuracy, precision, recall, F1-score showed model performs well in terms of predicting injury severity. Further Variable importance in terms of Sensitivity index explains impact of predictors on crash injury severity

4.2 Global Spatial Autocorrelation Analysis

Global Spatial Autocorrelation using Moran's I was performed on the integrated projected coordinates, yielding a z-score of 3.807630, p-value of 0.000140, Moran's index of 0.065825, and variance of 0.000306. The z-score exceeding 2.52 indicates the presence of high clusters, with a p-value < 0.01 suggesting less than a 1% probability that the observed clustered pattern is due to random chance. Moreover, assuming a random distribution of the analyzed characteristic across features, the null hypothesis was rejected. The Moran's Index value greater than 0 indicates a positive association. **Figure 4.1** illustrates the results and key values derived from the Global Moran's Index, depicting a normal distribution.

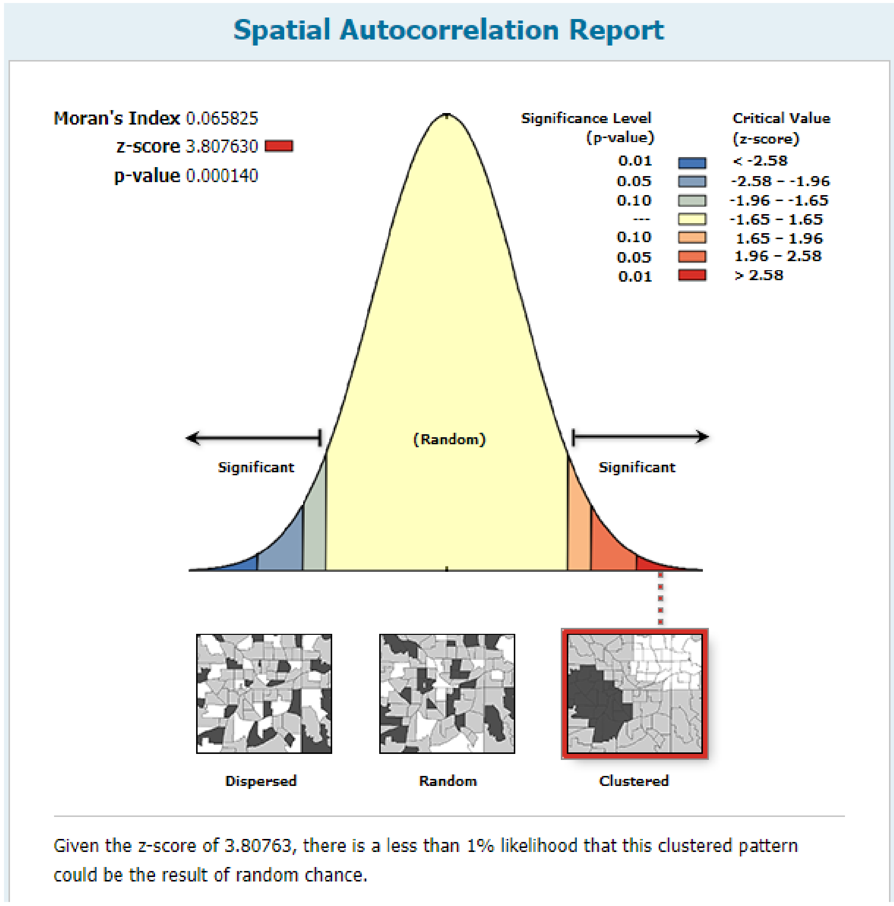

Figure 4.1: Global Moran's I Values

4.3 Incremental Spatial Autocorrelation Results

Incremental Spatial Autocorrelation was utilized to systematically investigate multiple distances and identify the most effective distance threshold for clustering of traffic crashes. The average distance was calculated using the "Calculate Distance Band from Neighbor Count", yielding an average distance of 172 meters. This distance was selected as the starting distance for Incremental Spatial Autocorrelation, with distance increments of 50 meters. Analysis revealed a peak clustering pattern at 272 meters (**Figure 4.2)**, with a z-score of 4.689852 and a p-value close to 0, indicating significant autocorrelation. **Table 4.1** provides detailed spatial autocorrelation results from various distances.

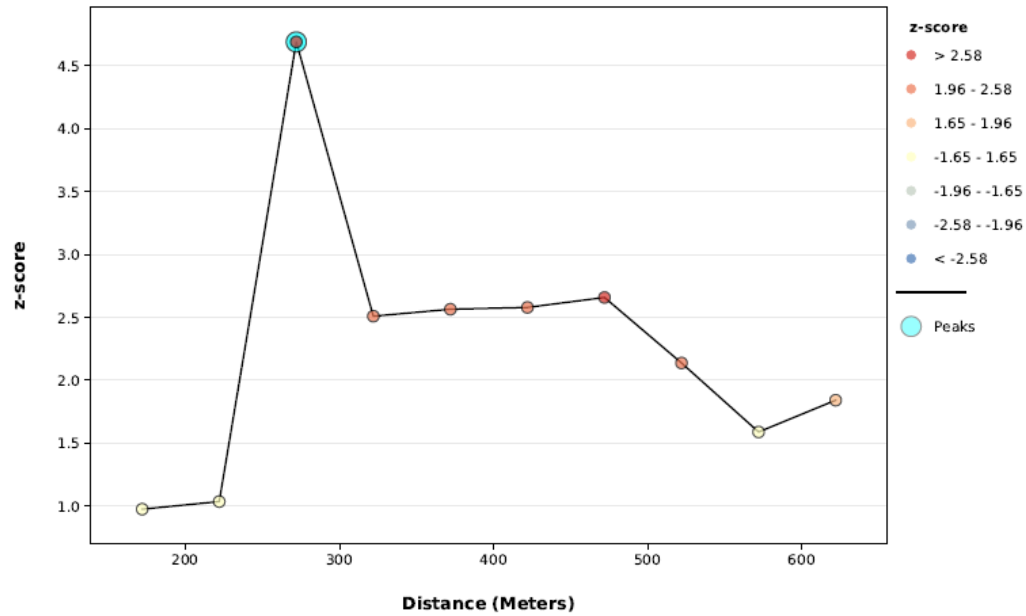

Figure 4.2: Plot of z value against distance increments

Table 4.1: ISA results for optimal clustering distance identification

Distance	Moran's Index	Expected Index	Variance	z-score	p-value
172.00	0.025207	-0.001042	0.000726	0.974474	0.39821
222.00	0.024978	-0.001003	0.000630	1.034927	0.300703
272.00*	**0.111472**	**-0.000965**	**0.000575**	**4.689852**	**0.000003**
322.00	0.055032	-0.000927	0.000497	2.509058	0.012105
372.00	0.052397	-0.000912	0.000432	2.564386	0.010336
422.00	0.049878	-0.000896	0.000388	2.578915	0.009911
472.00	0.048691	-0.000887	0.000348	2.657884	0.007863
522.00	0.037345	-0.000883	0.000320	2.136765	0.032617
572.00	0.025886	-0.000880	0.000284	1.588292	0.00222
622.00	0.028323	-0.000887	0.000252	1.840863	0.065642

*Peak (Distance value): 272.00 meters; z-score: 4.689852

4.4 Hotspots using Local Spatial Autocorrelation

For local Spatial Analysis, Hot spot analysis using Getis-Ord Gi* statistics was employed to identify significant hotspots based on Gi* scores. Weighted point data served as input for this analysis, with a threshold distance of 272 meters derived from the maximum peak of ISA. **Figure 4.3** illustrates the identified hotspots, where red, pink, and orange represent hotspots with 99%, 95%, and 90% significance levels, respectively. Red hotspots, predominantly located near bus stops such as Faizabad, Pirwidhai, and 26 Number bus stops, indicate concentrated areas of 3-MR activity due to frequent passenger pick-up and drop-off. These three locations are the major bus parking stops, where passengers come from within the city and other parts of the country and use 3-MRs for last-mile connectivity. One of the possible solutions to reduce the concentration of activities of 3-MRs in these locations is to provide feeder routes that should give access to people to important amenities within the city. Other significant hotspots are observed near commercial areas like Saddar, R.A. Bazar, Misrial Road, and Ayub Park where people access recreational and shopping amenities using 3-MRs. People use 3-MRs despite the safety risks associated with their unstable structure and risky driving behavior because of comparatively low fares to renting private vehicles. The other reason might be their smaller structure; 3-MR can easily approach the buses and shopping center entrances at their doorstep, providing easy access to people. Another important location of the hotspot is Racecourse Park on National Highway N-5, where the main parking of 3-MR is located. The other reason for this hotspot identification might be the presence of a driving license authority office, where mostly untrained drivers come for the driving test.

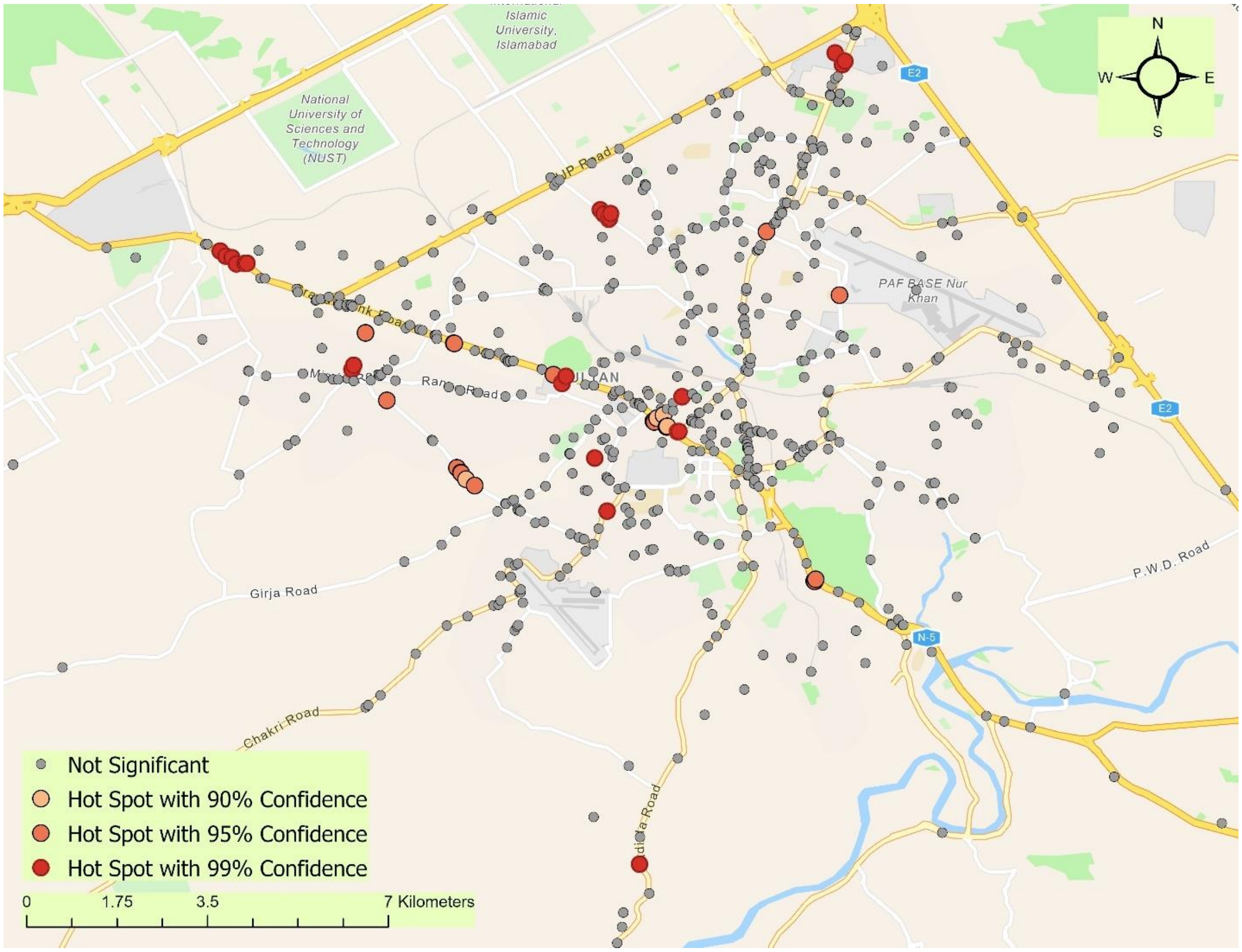

Figure 4.3: Hotspots Using Getis Ord Gi* Statistics

This local analysis of hotspots using Getis-Ord Gi* has pinpointed critical locations for crash occurrences involving 3-MR guiding authorities in allocating resources and implementing targeted interventions for safety measures. However, Getis-Ord Gi* does not account for the time factor in crashes. Therefore, Space-Time Cube for Emerging Hotspot Analysis was employed to incorporate the temporal dimension.

4.5 Emerging Hotspot Analysis using Space-Time Cube Analysis

Space-Time Cube was created using aggregated crash point data with a 1-month time step interval and a 1 km distance interval, converted into Network Common Data Format (netCDF). Emerging Hotspot Analysis was performed using this cube to classify each bin into hotspots based on Mann-Kendall test scores (Mann, 1945). **Figure 4.4** illustrates the space-time bins and resulting hotspot classifications.

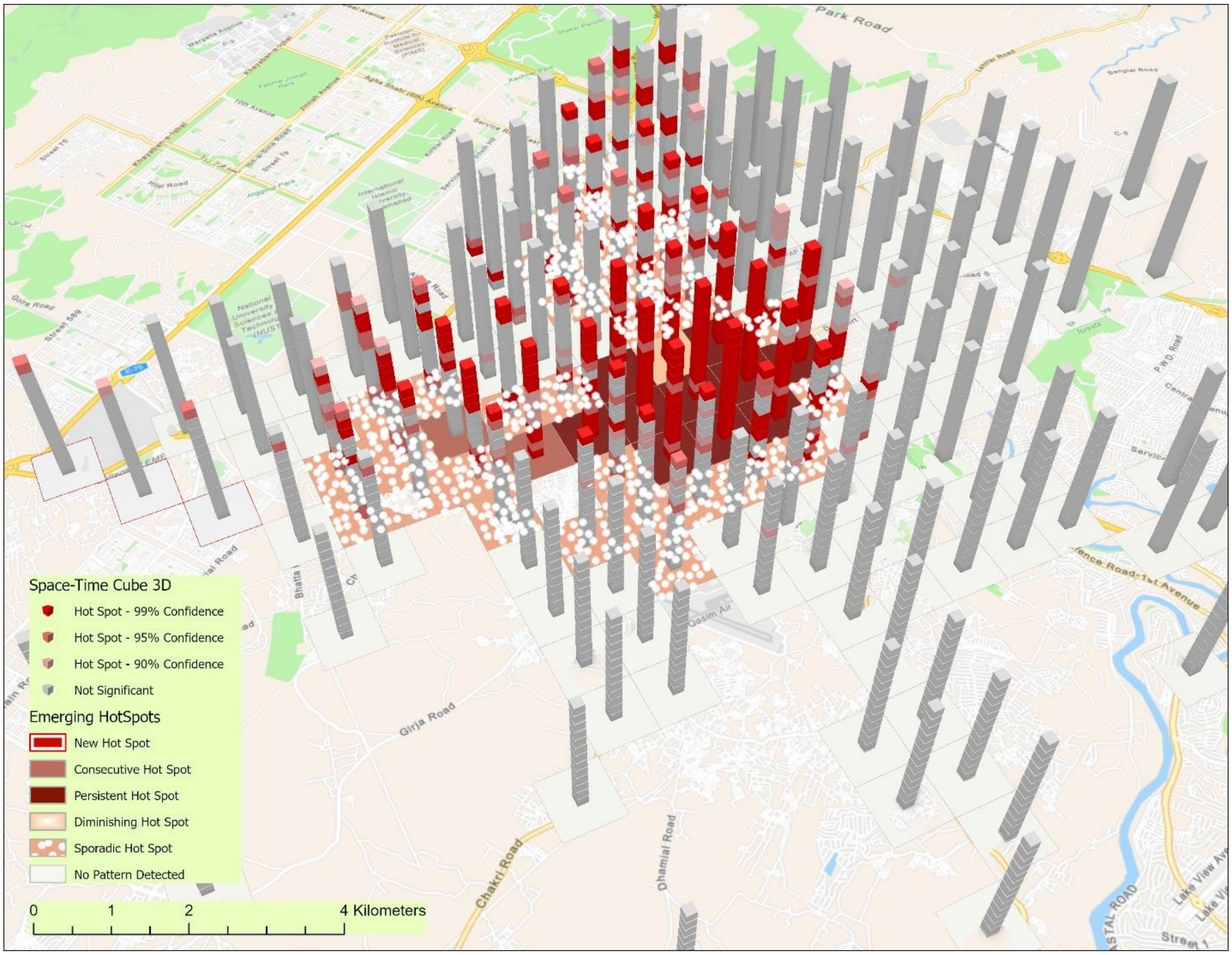

Figure 4.4: Space-Time Cube Bins

In total, 48 hotspots were identified, including 3 Consecutive, 10 Persistent, 1 Diminishing, 31 Sporadic, and 3 New hotspots as shown in **Figure 4.5**. New hotspots have been observed on a portion of Misrial Road near the National highway and motorway junction. Most bus stops are located at this place, which is probable cause for crashes as there is already congestion on the highway because of slow-moving vehicles, buses, public vehicles stopping and turning, and passenger movements.

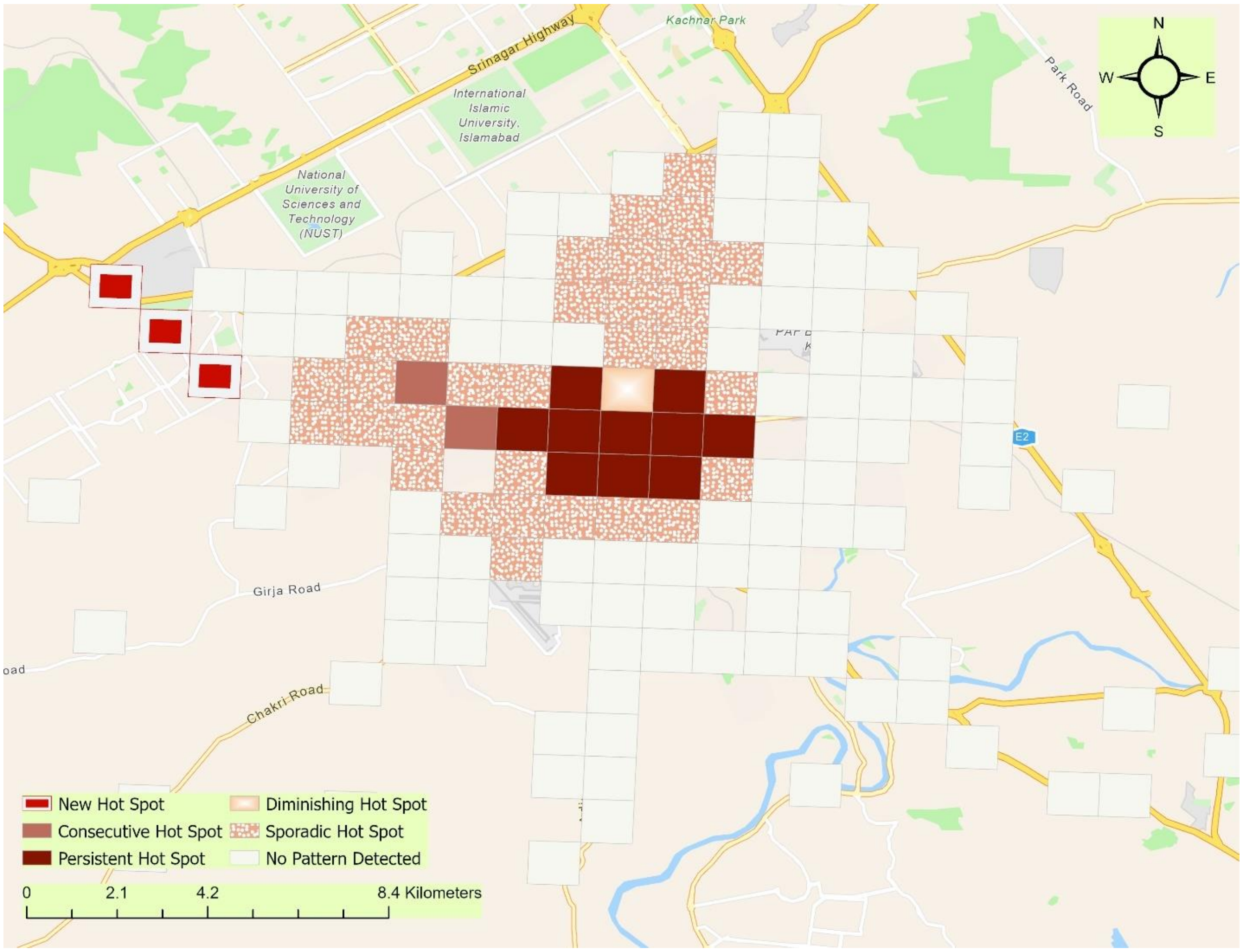

Figure 4.5: Emerging Hot Spots using Space time cube

Sporadic hotspots have been observed on Pirwadhai Bus stop, Faizabad Bus stop, Dhok Saiydan road, Dhamial Road, Abid Majeed Road, an intersection near CMH hospital, Range Road, Adiala Road, Fort Road, a portion of old airport road and Murree Road near Shamsabad. These locations suffer on-and-off phenomena regarding RTCs, which means specific time periods and road/traffic conditions lead to crash patterns in these areas. These locations can be summarized into four major categories, i.e., bus stops, proximity to hospitals and airports, and shopping centers. G.T. road (National Highway) from Racecourse Park to Kachehri Chowk intersection, Murree Road from Saddar to Liaqat Bagh Park, Railway Road, R.A. Bazar Road, Jhanda Chichi road, and Ammar Chowk intersection at old airport road have been identified as persistent hotspots. These areas lie where the most significant crash activity occurs, and these crashes have been hotspots throughout the analysis period without any increase or decrease in crash patterns. These locations typically represent government buildings, educational institutes, high commercial activity zones, and intersections. Areas of G.T. Road near Rafay Mall and Range Road near Byco petrol filling station have been consecutive hotspots, which means these areas have been significant hotspots recently and have not experienced hotspots before. The possible reason for this could be significant road infrastructural changes, such as providing protected U-turns and replacing the intersections. With the provision of these protected U-turns, the drivers need to frequently change lanes, which may lead to an increased frequency of crashes. Also, these locations are associated with high-density residential movement activities making the area more vulnerable. Some portions of the railway road near City Saddar Road have been associated with diminishing hotspots as these areas also experienced RTCs, but the clustering activity at these locations declined through the analysis period. The reason for this diminishing hotspot could be converting an existing two-way road to one-way during the analysis period, possibly reducing the traffic conflicts.

The comprehensive hotspot analysis of 3-MR crashes identified areas with sustained crash patterns and also revealed emerging trends and specific time-dependent crash occurrences. Such insights are pivotal for informing policy decisions aimed at improving road safety through targeted resource allocation and intervention strategies. The advancement represented by these analyses in the Rawalpindi district underscores their potential for guiding future transportation policies and research endeavors. By leveraging spatial and spatiotemporal data analytics, policymakers can implement evidence-based interventions that address local and temporal variations in crash occurrences, ultimately enhancing overall road safety and fostering sustainable urban mobility.

4.6 Spearman Correlation

Spearman correlation was performed to analyze the correlation and multicollinearities between variables. Spearman correlation uses monotonic functions to assess the correlation relationships between features. High correlation between independent variables indicates multicollinearity (Abdullah et al., 2023) and such variables give redundant information which makes it difficult to determine unique contribution of each variable. Multicollinearity of independent variables also impacts parameter estimates in case of regression models i.e. by increasing such variables in data even on a small scale can cause huge changes in regression coefficients. In case of neural networks, they are somewhat insensitive to multicollinearity (De Veaux & Ungar, 1994) but still highly correlated variables may cause overfitting problems. **Figure 4.6** displays the heatmap matrix of spearman correlation coefficient values for each variable which range between -1 to +1. Dark blue color shows strong negative correlation and values approaching -1 while red color shows strong positive correlation with values near +1, and lighter color denotes weaker correlations. Highly correlated factors were removed or only one was kept like daily maximum, daily minimum and daily average temperature were highly correlated so only average temperature was used among predictor variables. Similarly light conditions day/night, time am/pm, and hourly time classes were also highly correlated so only hourly time classes were used for ANN model. Day of week and month of year were removed and weekend/weekday and season of year were used among predictor variables.

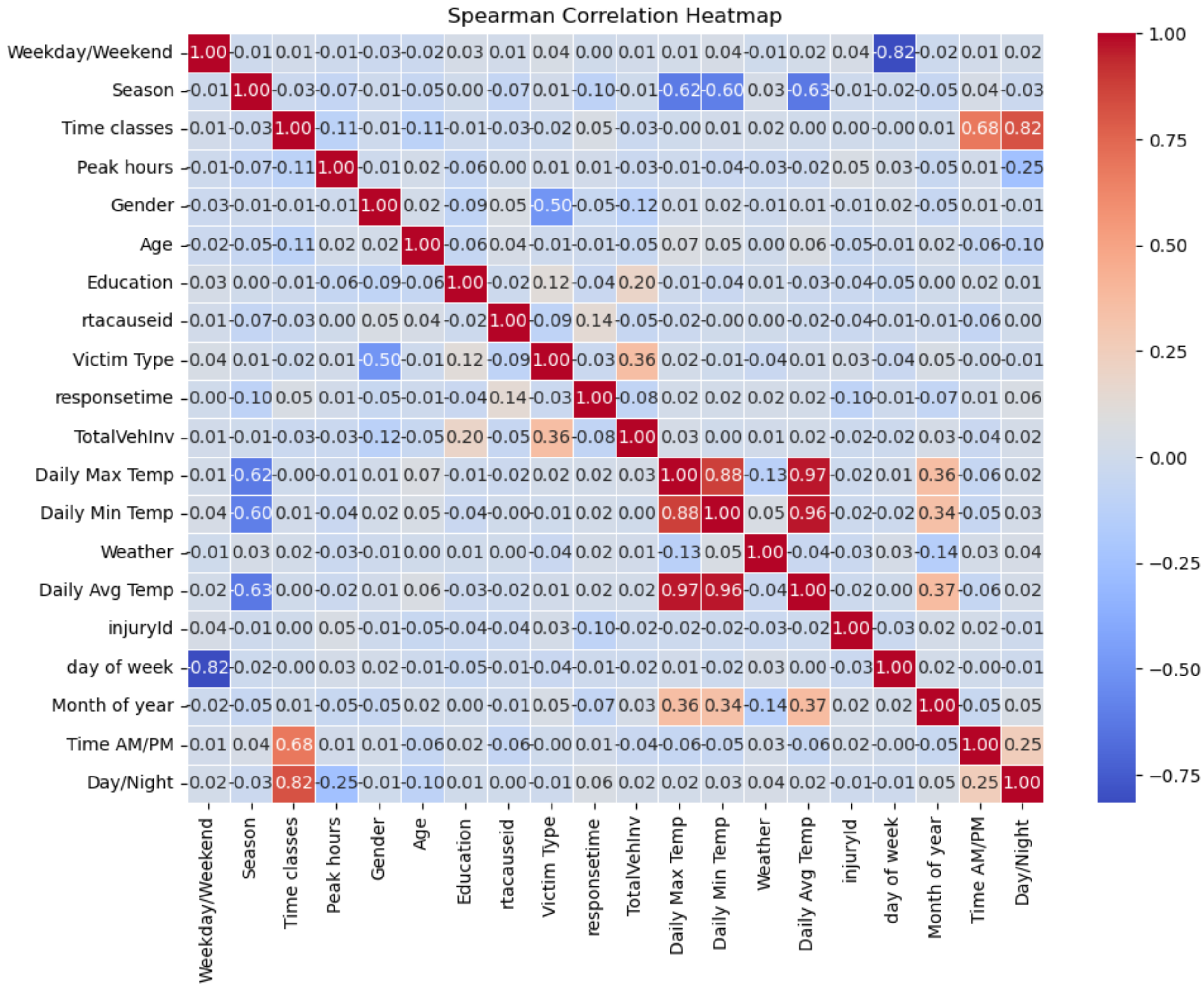

Figure 4.6: Spearman correlation among variables

4.7 Artificial Neural Network

Data used to make prediction model using ANN was over sampled through Synthetic Minority Oversampling Technique (SMOTE) which increased the number of samples of minority class from 298 to 1122, making it similar to majority class. So, a total of 2244 samples were used in developing model. To develop a prediction model using Artificial Neural Networks we used training dataset of 70% from overall data, and model accurately predicted 74.4% of values with Sum of Squares Error of 269 and for testing data set 20% of overall data was used and model accurately predicted 73.3% of values with Sum of Squares Error of 86. While for validation 10% of data was used and model accurately predicted 73.9% of values.

4.7.1 Performance Metrics

In order to check the performance of ANN, several performance metrics were analyzed. Confusion matrix for training, testing and validation data is shown in **Table 4.2**. In the training set, the confusion matrix reveals that out of the observed severe injury cases, 587 were correctly predicted as severe, while 173 were incorrectly classified as minor, resulting in a correct prediction rate of 77.2%. Similarly, for observed minor injury cases, 545 were accurately predicted as minor, with 217 misclassified as severe, yielding a correct prediction rate of 71.5%. Overall, the model for achieved an accuracy of 74.4% in the training set, with 52.8% of predictions being severe and 47.2% being minor.

The confusion matrix in the testing set exhibits comparable performance to that in the training set. The accuracy of the predictions was 71.6%, with 164 serious injury cases accurately identified as severe and 65 cases incorrectly marked as minor. The accuracy percentage of the predictions was 75%, with 180 minor injury cases correctly identified as minor and 60 instances incorrectly labelled as severe. The model achieved an overall accuracy of 73.3% in the testing set, with 47.8% of predictions classified as severe and 52.2% classified as minor. Validation of the model also performed better with confusion matrix showing comparable performance. For severe Injuries model correctly identified 97 injuries as Severe and misclassified 36 Severe injuries as Minor with 72.9% correct percentage. For classification of Minor injuries model correctly identified 90 injuries as Minor and misclassified 30 as Severe Injuries with correct percentage of 73.9%.

Table 4.2: Confusion Matrix values

Sample	Observed	Predicted		
		Severe	Minor	Percent Correct
Training	Severe	587	173	77.20%
	Minor	217	545	71.50%
	Data Distribution	52.80%	47.20%	74.40%
Testing	Severe	164	65	71.60%
	Minor	60	180	75.00%
	Data Distribution	47.80%	52.20%	73.30%
Validation	Severe	97	36	72.90%
	Minor	30	90	75.00%
	Data Distribution	50.20%	49.80%	73.90%

Table 4.3: Performance Metrics Parameters

	Precision	Accuracy	Recall	F1-Score	Specificity
Training	77.24%	74.38%	73.01%	75.06%	75.91%
Testing	71.62%	73.35%	73.21%	72.41%	73.47%
Validation	72.93%	73.91%	76.38%	74.62%	71.43%
Overall	76.08%	74.45%	73.70%	74.87%	75.26%

Performance Metrics are shown in **Table 4.3.** Precision exhibited a comparable level between the training, testing and validation sets, with values of 77.24%, 71.62% and 72.93%respectively. This indicates a notable level of reliability in the model's predictions. The training set achieved an accuracy rate of 74.38%, the testing set yielded an accuracy rate of 73.35% and validation set had an accuracy of 73.91%. Model had an overall greater precision and accuracy for training, but model also achieved good to comparable performance while testing and validating data.

The assessment of recall (sensitivity) and specificity scores demonstrated the model's capacity to precisely detect both true positive and true negative instances. The recall scores for severe injury cases in the training set were calculated to be 73%, while for minor injury cases it was 75.9%. Similarly, the specificity scores for severe and minor injury categories were 75.9% and 73% respectively. In the testing set, the recall scores for severe injury cases were 73.2% and for minor injury cases were 73.5%, while specificity scores also showed similar results of 73.5 for severe injury and 73.2%. The recall score defining the Severe injuries in the validation set were 76.4% and 71.4% for Minor injury and specificity scores were 71.4% and 76.4%. F1-Score, which combines precision with the recall rate also presented a fairly good predictive competency in the Training, Testing and Validating sets, with scores of 75. 06%, 72. 41% and 71. 43%. The balanced performance of the ANN model, as indicated by both precision and recall, reinforces its dependability in forecasting the severity of traffic crash injuries. Results also demonstrated effectiveness of the proposed model in detecting severe and minor injuries and eliminating the false positives and false negatives.

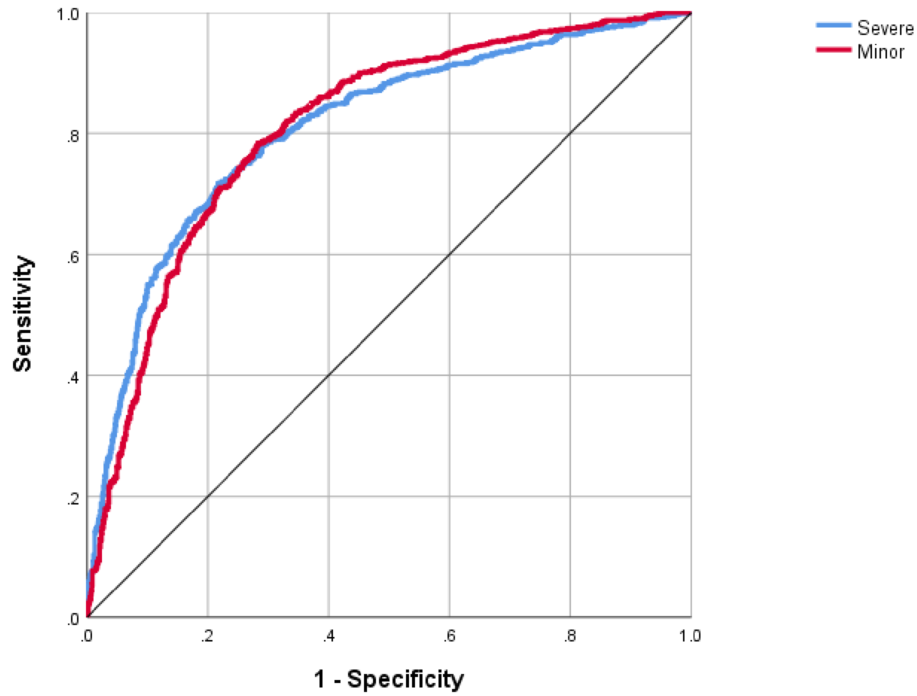

Figure 4.7: ROC Curve

The Receiver Operating Characteristic (ROC) curve **Figure 4.7** illustrates the trade-off between sensitivity and specificity. The data demonstrates that the model repeatedly outperformed random chance, as evidenced by its significant deviation from the diagonal line. The model's ability to distinguish between severe and minor cases was confirmed by the *Area Under the Curve* (AUC) values, which were 0.809 for both categories.

ANN model's exceptional predictive abilities in assessing the severity of traffic crash injuries are evident from the consistently high AUC values, precision, recall, specificity, and F1-Score metrics observed throughout both the training and testing datasets. The results emphasize the model's capacity to enhance decision-making processes regarding resource allocation and intervention strategies in the context of traffic events.

4.7.2 Synaptic Weights through Neural Network

Synaptic weights from input layer to hidden layer and from hidden layer to output layer are given in a form of table in **Appendix A**. Response time gives positive weights to (H(1:1), H(1:2), H(1:3), H(1:5), H(1:8)) hidden layers and negative to (H(1:4), H(1:6), H(1:7)), and synapses (H(1:1), H(1:2)) contribute positively to severe injury (Injury=1) and (H(1:4), H(1:6), H(1:7)) contribute negatively to severe injury in output layer. The weights for response time range from -0.736 H(1:4) to 0.563 H(1:2). Average temperature gives significant weights to (H(1:1), H(1:2), H(1:3), H(1:5) Higher average temperatures increase the activation of neurons which in turn increases the probability of severe injuries and has negative influence on minor injuries. The weights for Number of Victims range from -0.558 H(1:2) to 1.204 H(1:4). From all the weight inputs to layers, more victims lead to slight increase in activation of final output of severe injuries. Weights for Day range from -0.646 H(1:7) to 0.864 H(1:2). Season contributed positive weights to H(1:2) (0.550), H(1:3) (0.920), H(1:5) (0.473), H(1:8) (0.307) and Negative weights H(1:1) (-0.474), H(1:4) (0.065), H(1:6) (-0.783), H(1:7) (-0.235). Weights from Season have positive influence on Minor Injury, except H(1:5) which positively influence Severe injury. Time gives positive weight to 3 neurons with highest being 0.400 to H(1:7) and negative weight to 5 neurons with lowest -0.581 to H(1:5). The weights for Education input range from -1.133 H(1:5) to 0.488 H(1:3). Higher education reduces severe injuries by decreasing H(1:5), while H(1:1), (H(1:6), H(1:7), H(1:8), leads to more severe injuries which shows education has impact on both categories. Peak hours activate neurons H(1:1), H(1:3), H(1:4), H(1:6), and H(1:7), while deactivating H(1:2), H(1:5), and H(1:8). Neurons H(1:2), H(1:3), H(1:4), H(1:6), H(1:7), and H(1:8) favor minor injuries, whereas H(1:1) and H(1:5) lean towards severe injuries. Gender differences activate neurons H(1:1), H(1:3), H(1:4), and H(1:6), while deactivating H(1:2), H(1:5), H(1:7), and H(1:8) indicating certain type of gender activates respective injury output. For RTA cause the weights range from -0.501 H(1:5) to 0.874 H(1:2). RTA cause have mixed impact, positive weights indicated certain causes contribute to severe injuries while other to minor injuries. Weights by victim type range from -.632 H(1:1) to 2.103 H(1:6). Age had the most positive influence on H(1:2) of 0.916 and the most negative influence on H(1:1) of -0.571.

The bias weights for the hidden layer nodes H(1:1) to H(1:8) are 0.251, -0.148, 1.373, -0.048, -0.073, -0.065, -0.520, and 0.230 respectively.The bias weights for the output layer nodes [Injury=1] and [Injury=2] are 0.132 and -0.127 respectively.

4.7.3 Pseudo Probability of Injury Classes

Use of sigmoid function for output layer results in generation of results in the form of 0 and 1. So from these values probabilities of classification of cases can be interpreted and are known as pseudo probability. By evaluating the pseudo probability, we can analyze the model's confidence in predicting the correct cases. The **Figure 4.8** displays the predicted pseudo-probabilities of both injury classes Severe and Minor. For Severe injuries cases classified as Severe, the blue bars indicate a high pseudo-probability of correct classification, reaching up to 0.81 and for minor injuries cases classified as Minor, the red bars show a high pseudo-probability of correct classification, also reaching about 0.8. Model effectively differentiates between Severe and Minor injuries, with a higher likelihood of accurate classification for each respective category. This suggests that the model has learned the patterns in the data well and can be a reliable tool for predicting injury severity in traffic crashes.

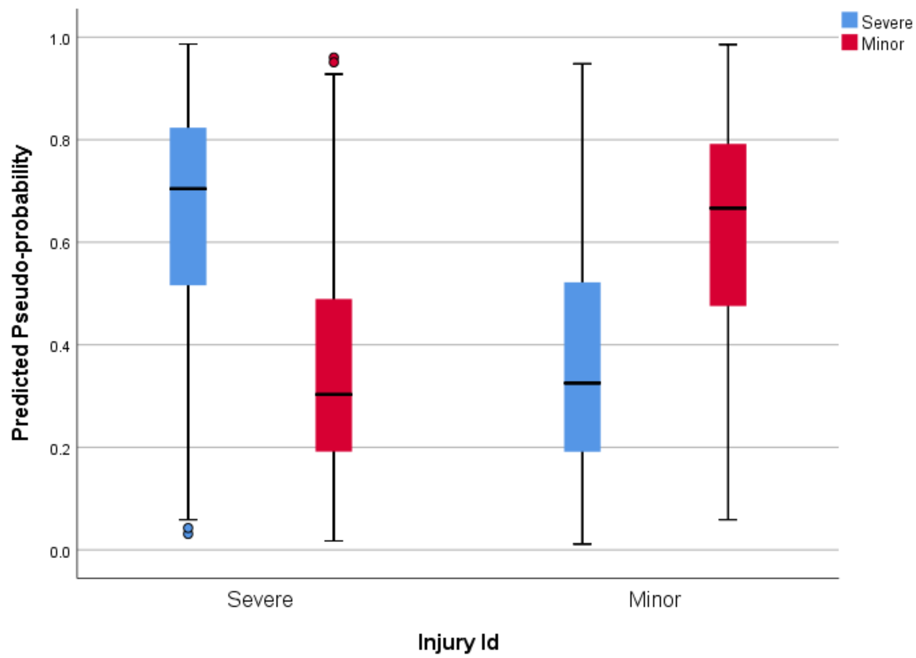

Figure 4.8: Predicted Pseudo Probability of injury classes

4.7.4 Variable Importance using Sensitivity Index Analysis

Sensitivity analysis was also carried out by the ANN model. Variance-based methodology was employed to assess the significance of risk factors for crash severity in SPSS. Independent Variable analysis uses variance-based technique in SPSS in order to rank input variables based on their contribution to injury severity prediction during model training. The sensitivity measure S_i determines the relative significance of the variables for every given set of interactions and non-orthogonality between them (Amiri et al., 2021).

$$S_i = \frac{V_i}{V(Y)} = \frac{V(E(Y/X_i))}{V(Y)}$$

Where S_i is the Sensitivity Index, V(Y) is the unconditional output variance, E refers to an integral over X_i, and variance operator V denotes a further integral over X_i (Saltelli et al., 2004).

Table 4.4: Sensitivity Index Values

Variables	Importance	Relative Contribution
Response time	0.108	100.0 %
RTA cause	0.09	83.3 %
Average temp	0.089	82.4 %
Gender	0.087	80.6 %
Age	0.081	75.0 %
Victim Type	0.08	74.1 %
Number of Victims	0.079	73.1 %
Peak hours	0.07	64.8 %
Season	0.065	60.2 %
Education	0.059	54.6 %
Time	0.057	52.8 %
Weather	0.056	51.9 %
Day	0.052	48.1 %
Vehicles Involved	0.027	25.0 %

Table 4.4 shows S_i importance of each variable and **Figure 4.9** shows normalized importance of factors on crash injury. Response time is the highest contributor, response time itself cannot predict the injury outcome but if timely aid can be delivered many fatal

injuries can be avoided which occur as a result of blood loss or such conditions which if treated within time can save a life. Probability of severe injuries significantly increases if the crash occurs either due to over speeding or distracted driving. 55.4% of severe injuries occurred because of over speeding at the time of crash while reckless driving caused 38.5% severe injuries. Indicating over speeding and careless driving behavior needs to be addressed in order to reduce severe injury crashes. Temperature of the day and Gender also show significant contribution which can be because most injuries and crashes involved males and resulted in more injuries. Age group 18-30 years increases the likelihood of crashes leading to severe injuries Specific age group of 18-30 years contributed about 40% of severe cases to injury. Drivers experienced 56% of severe injuries, 36% severe injuries were encountered by passengers. When drivers are in young age group and cause of accident is over speeding and careless driving it increases the chance of severe injuries. There is 50% chance of severe injuries if both of these conditions occur together.

Peak hours lead to a smaller number of severe injuries then off-peak that is due to slower traffic speeds during peak hours despite higher traffic volumes. Spring season experienced a highest number of severe injuries. Lack of education also plays important role in predicting injury severity, as drivers who had little to no schooling caused 51% of severe crashes, while drivers which had education level up to matric caused 30% of severe crashes. 12 pm noon to evening 6 pm time contributed to most of severe injury cases. Weekdays contributed more to injury severity, about twice the severe injuries occurred on weekdays as compared to severe injuries on weekdays.

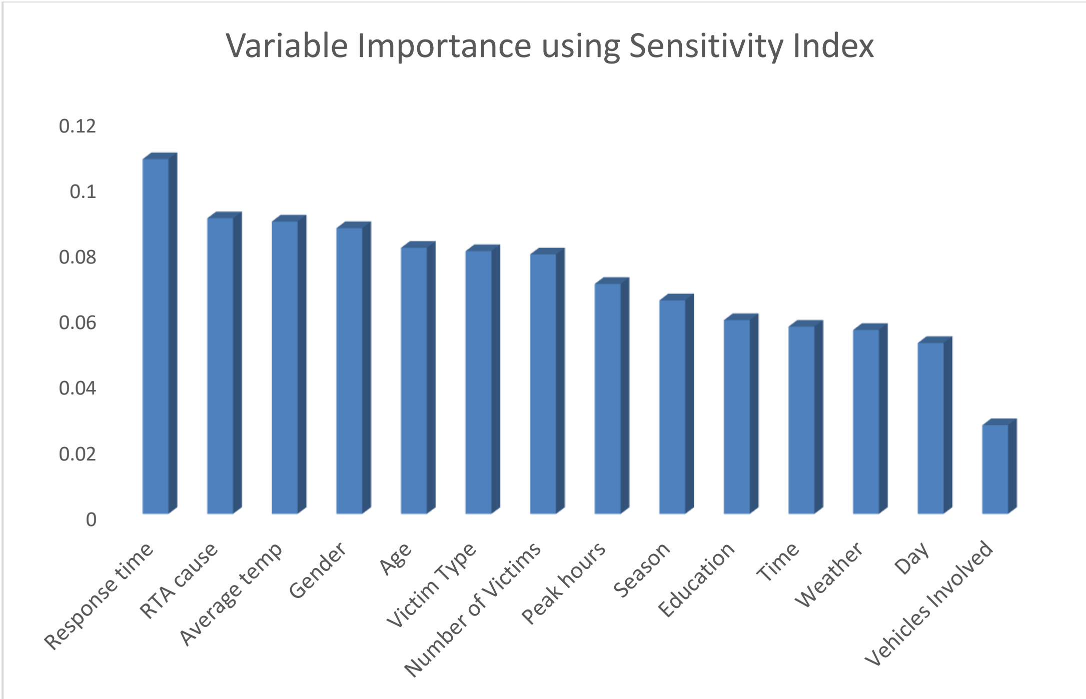

Figure 4.9: Variable Importance Ranking using Sensitivity Index

4.7 Summary of Results

- As the data is acquired from Rescue 1122, which is an emergency response service, so response time was also taken into account when predicting injury severity, and response time came out to be of highest importance, response time doesn't predict injury severity before crash conditions, but if response time is reduced it may significantly reduce those severe injuries which happens because of no timely aid delivered, like excessive blood loss etc.
- Chances of severe injury increases if gender is male and crash is caused due to over speeding in sunny weather and driver has little to no education.
- Young age between 18-30 years, time of day from noon to evening, and weekdays also increase the probability of severe injury crashes.

- Crash with another vehicle i.e. when two vehicles are involved in accident, it leads to severe injuries.

CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

5.1 Summary

This study carried out a comprehensive analysis of traffic crashes of 3-MR. Data of 3-MR traffic crashes from January 2022 to April 2023 was used in analysis. To facilitate analysis in ArcGIS location of crashes were geocoded and converted into coordinates in form of latitude and longitude. Analysis of traffic crash hotspots was done in ArcGIS using spatial and temporal dimension. Further prediction modeling using ANN was done in order to predict injury severity of 3-MR traffic crashes. Both techniques discovered different aspects of 3-MR traffic crashes which are discussed in next sections.

5.2 Conclusions based on Hotspot Analysis

This study applied various spatial and spatiotemporal analyses to 3-MR crash data in Rawalpindi district. Global Spatial Autocorrelation using Moran's I revealed significant clustering of crash points, with a z-score of 3.807630, p-value of 0.000140, and Moran's index of 0.065825, indicating a positive association and rejecting the null hypothesis of random distribution. In the next step, ISA identified 272 meters as the optimal distance threshold for clustering, showing a peak z-score of 4.689852 and a p-value close to 0. Local Spatial Analysis through Getis-Ord Gi* Hot Spot was also performed to identify the 3-MR crash hotspots by location. To incorporate the time dimension, Emerging Hotspot Analysis using Space-Time Cube was performed, revealing 48 hotspots categorized into Consecutive, Persistent, Diminishing, Sporadic, and New hotspots. This analysis highlighted areas with sustained crash patterns, emerging trends, and specific time-dependent crash occurrences. The spatial and spatiotemporal analyses of traffic crash data in Rawalpindi district identified critical clusters near bus stops, commercial areas, hospitals and airports, intersections, and high-density residential units. These findings underscore the need for targeted interventions focusing on these hotspots, ensuring effective resource allocation and evidence-based policies to enhance road safety and urban mobility.

5.3 Conclusions based on Prediction Modeling using Artificial Neural Network

Three wheeled motorized rickshaws are crucial part of traffic in urban cities of Pakistan, which have been neglected in terms of road safety analysis. This study also analyzed the impact of different factors on injury severity of 3-MR crashes using Artificial Neural Network (ANN) with back propagation algorithm. Synthetic Minority Oversampling Technique (SMOTE) was also employed to deal with class imbalance between Severe and Minor injury severity. Severe injury class was underrepresented in data which led to ANN model overfitting and model only favored Minor class in its prediction. SMOTE using k-nearest neighbors created synthetic samples for underrepresented class, and data from SMOTE was used as input for ANN Modelling. Various performance metrics were analyzed which showed model achieved satisfactory performance. Model achieved an accuracy of 74.5% and Precision of 75.1% with true positive ratio of 73.7% and false positive ratio of 75.3%. Receiver Operating Characteristics (ROC) curve was also analyzed, area under the curve value of 0.809 for both severe and minor injury classes depicted ability of model to distinguish between both classes. Further Spearman correlation and variable importance using sensitivity index were also evaluated to understand the correlation between input variables and impact of each variable in prediction of target variable. Further variable importance evaluated contribution of predictors for injury severity in ANN model using Sensitivity index. Response time, cause of crash, temperature and gender had all significant impact on predicting injury severity showed. Mentioned variables also had normalized importance of above 80%. Higher response times, over speeding and careless driving behaviors, and high temperature increase the risk of severe injury crash. Male gender, time from 12pm-6pm, off peak periods and weekdays also have effect on injury severity. Age group of 18-30 years, little to no education and impact with other vehicle also amplified the probability of severe injuries. Overall, ANN model showed satisfactory results and SMOTE resolved the problem of overfitting and established a balance between both classes of target variable.

5.4 Future Recommendations

This study analyzed the crash hotspots of traffic crashes using spatial and temporal distribution of 3-MR traffic crashes. As significant hotspots have been identified near bus stops, commercial areas, hospitals, airports, intersections, and high-density residential units.

- Measure such as separate stops for 3-MR on bus stops and commercial areas should be taken into account so that pick and drop of passengers for 3-MR can be facilitated as such areas are prone to these type of crashes where pulling off or getting into lane can be quite crucial for these types of vehicles.
- Strict law enforcement should be done on Intersections and U-turns, where violation of signals and moving wrong way is common for 3-MR.
- Mandatory speed limits near residential areas should be enforced. Overall, there are rules for speed limits when driving through residential areas, school zones, hospital etc. but need of the hour is to enforce those laws.

- Emergency response services should make sure for timely response in case of traffic crash. In case of rush hours such services need to have backup plan such that to avoid high traffic areas. Places where crash hotspots have been identified should be prioritized in case of resource allocation for immediate emergency response.
- As majority of drivers of 3-MR lack proper education and schooling, so license issuing authorities like traffic police should make a proper curriculum regarding traffic safety and rules of driving, and proper classes to be taken for anyone who apply for a license. This will not only educate drivers but also create awareness regarding traffic safety.
- 3-MR Traffic could be diverted from some routes as a remedial measure where high concentration of 3-MR crash hotspots exists, and if it proves to impactful then it should be implemented further. Facilities being limited to one way traffic only has proven to improve traffic safety in past.
- Dedicated lanes should be given for 3-MR as well as two-wheelers on high-speed roads, as impact with other fast-moving vehicle severely worsens the crash injury outcome of 3-MR and two wheelers.

Future research can incorporate hotspot analysis techniques for 3-MR crashes by considering injury types (severe, minor, etc.). Similarly, hotspot analysis can be done for 3-MR crashes by nature of the crash (head-on, rear-end, side-swap, and roll-over) and crash with vehicle mode (two-wheelers, cars, trucks, etc.).

REFERENCES

Abdel-Aty, M. (2003). Analysis of driver injury severity levels at multiple locations using ordered probit models. Journal of Safety Research, 34(5), 597–603. https://doi.org/10.1016/j.jsr.2003.05.009

Ahmed, A., Khan, B. A., Khurshid, M. B., Khan, M. B., & Waheed, A. (2016). Estimating national road crash fatalities using aggregate data. International Journal of Injury Control and Safety Promotion, 23(3), 249–254. https://doi.org/10.1080/17457300.2014.992352

Ali, Q., Yaseen, M. R., & Khan, M. T. I. (2020). The impact of temperature, rainfall, and health worker density index on road traffic fatalities in Pakistan. Environmental Science and Pollution Research, 27(16), 19510–19529. https://doi.org/10.1007/s11356-020-08233-1

Alkaabi, K. (2023). Identification of hotspot areas for traffic accidents and analyzing drivers' behaviors and road accidents. Transportation Research Interdisciplinary Perspectives, 22. https://doi.org/10.1016/j.trip.2023.100929

Alkheder, S., Taamneh, M., & Taamneh, S. (2017). Severity Prediction of Traffic Accident Using an Artificial Neural Network. Journal of Forecasting, 36(1), 100–108. https://doi.org/10.1002/for.2425

Amarasingha, N. (2015). Characteristics of Three-Wheeler Crashes. The 5th International Research Symposium on Engineering Advancements, SAITM, Colombo, Sri Lanka. https://www.researchgate.net/publication/286473386

Amin, S. (2020). Backpropagation – Artificial Neural Network (BP-ANN): Understanding gender characteristics of older driver accidents in West Midlands of United Kingdom. Safety Science, 122. https://doi.org/10.1016/j.ssci.2019.104539

Amiri, A. M., Naderi, K., Cooper, J. F., & Nadimi, N. (2021). Evaluating the impact of socio-economic contributing factors of cities in California on their traffic safety condition. Journal of Transport and Health, 20. https://doi.org/10.1016/j.jth.2021.101010

Anselin, L. (1995). Local Indicators of Spatial Association—LISA. Geographical Analysis, 27(2), 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Arain, M. A., Ahsan, N., Aatif, W. K., & Ahmed, R. A. (2017). Demographic Determinants of Physical Injuries in Rawalpindi: Role of Rescue 1122. JIMDC, 6(3), 148–152.

Awan, W. S., Ahmed, A., Bilqees, U., Malik, U., Naseer, M. K., & Qamar, M. S. I. (2020). Injury patterns in patients with QINGQI motorcycle rickshaw related road traffic accidents presenting at accident and emergency department of Mayo Hospital, Lahore. The Professional Medical Journal, 27(07), 1476–1481. https://doi.org/10.29309/tpmj/2020.27.07.4438

Bader-El-Den, M., Teitei, E., & Perry, T. (2019). Biased Random Forest for Dealing with the Class Imbalance Problem. IEEE Transactions on Neural Networks and Learning Systems, 30(7), 2163–2172. https://doi.org/10.1109/TNNLS.2018.2878400

Batool, I., Hussain, G., Kanwal, N., & Abid, M. (2018). Identifying the factors behind fatal and non-fatal road crashes: a case study of Lahore, Pakistan. International Journal of Injury Control and Safety Promotion, 25(4), 401–407. https://doi.org/10.1080/17457300.2018.1456466

Bishai, D., Quresh, A., James, P., & Ghaffar, A. (2006). National road casualties and economic development. Health Economics, 15(1), 65–81. https://doi.org/10.1002/hec.1020

Bisht, L. S., & Tiwari, G. (2023). Identification of road traffic crashes hotspots on an intercity expressway in India using geospatial techniques. IATSS Research, 47(3), 349–356. https://doi.org/10.1016/j.iatssr.2023.07.003

Bokaba, T., Doorsamy, W., & Paul, B. S. (2022). Comparative Study of Machine Learning Classifiers for Modelling Road Traffic Accidents. Applied Sciences (Switzerland), 12(2). https://doi.org/10.3390/app12020828

Budzyński, M., Kustra, W., Okraszewska, R., Jamroz, K., & Pyrchla, J. (2018). The use of GIS tools for road infrastructure safety management. E3S Web of Conferences, 26. https://doi.org/10.1051/e3sconf/20182600009

Butt, A., Ahmad, S., Shabbir, R., & Erum, S. (2017). GIS based surveillance of road traffic accidents (RTA) risk for Rawalpindi city: a geo-statistical approach. Kuwait J. Sci, 44(4), 129–134.

Çela, L., Shiode, S., & Lipovac, K. (2013). Integrating GIS and Spatial Analytical Techniques in an Analysis of Road Traffic Accidents In Serbia. International Journal for Traffic and Transport Engineering, 3(1), 1–15. https://doi.org/10.7708/ijtte.2013.3(1).01

Chakraborty, A., & Maitra, B. (2024). Riding behaviours of motorcyclists in mixed traffic condition and its association with crash risk. Case Studies on Transport Policy, 16. https://doi.org/10.1016/j.cstp.2024.101177

Cheng, Z., Zu, Z., & Lu, J. (2019). Traffic crash evolution characteristic analysis and spatiotemporal hotspot identification of urban road intersections. Sustainability (Switzerland), 11(1). https://doi.org/10.3390/su11010160

Constant, A., Salmi, L. R., Lafont, S., Chiron, M., & Lagarde, E. (2008). The recent dramatic decline in road mortality in France: How drivers' attitudes towards road traffic safety changed between 2001 and 2004 in the Gazel cohort. Health Education Research, 23(5), 848–858. https://doi.org/10.1093/her/cym072

Dandona, R., Anil Kumar, G., & Dandona, L. (2005). Traffic law enforcement in Hyderabad, India. International Journal of Injury Control and Safety Promotion, 12(3), 167–176. https://doi.org/10.1080/17457300500088840

Deshpande, N., Chanda, I., & Arkatkar, S. S. (2011). Accident Mapping And Analysis Using Geographical Information Systems. International Journal of Earth Sciences and Engineering, 04(SPL), 20410227.

Eskandari Torbaghan, M., Sasidharan, M., Reardon, L., & Muchanga-Hvelplund, L. C. W. (2022). Understanding the potential of emerging digital technologies for improving road safety. Accident Analysis and Prevention, 166. https://doi.org/10.1016/j.aap.2021.106543

ESRI. (2023a). Emerging Hot Spot Analysis (Space Time Pattern Mining). https://pro.arcgis.com/en/pro-app/2.8/tool-reference/space-time-pattern-mining/emerginghotspots.htm

ESRI. (2023b). How Emerging Hot Spot Analysis works—ArcGIS Pro | Documentation. https://pro.arcgis.com/en/pro-app/2.8/tool-reference/space-time-pattern-mining/learnmoreemerging.htm

ESRI. (2023c). Incremental Spatial Autocorrelation (Spatial Statistics). https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/incremental-spatial-autocorrelation.htm

Gadepalli, R., Tiwari, G., & Bolia, N. (2020). Role of user's socio-economic and travel characteristics in mode choice between city bus and informal transit services: Lessons from household surveys in Visakhapatnam, India. Journal of Transport Geography, 88. https://doi.org/10.1016/j.jtrangeo.2018.08.017

García de Soto, B., Bumbacher, A., Deublein, M., & Adey, B. T. (2018). Predicting road traffic accidents using artificial neural network models. Infrastructure Asset Management, 5(4), 132–144. https://doi.org/10.1680/jinam.17.00028

Getis, A., & Ord, J. K. (1992). The Analysis of Spatial Association by Use of Distance Statistics. Geographical Analysis, 24(3), 189–206. https://doi.org/10.1111/j.1538-4632.1992.tb00261.x

Glen, S. (2023). Moran's I: Definition, Examples. https://www.statisticshowto.com/morans-i/

Goel, R. (2018). Modelling of road traffic fatalities in India. Accident Analysis and Prevention, 112, 105–115. https://doi.org/10.1016/j.aap.2017.12.019

Gudes, O., Varhol, R., Sun, Q. (Chayn), & Meuleners, L. (2017). Investigating articulated heavy-vehicle crashes in western Australia using a spatial approach. Accident Analysis and Prevention, 106, 243–253. https://doi.org/10.1016/j.aap.2017.05.026

Hambissa, D., Nallamothu, R. B., & Andarge, T. (2022). Analysis of Three-Wheeler Vehicle Structure at the Event of Side Pole Crash. Journal of Engineering (United Kingdom), 2022. https://doi.org/10.1155/2022/5490585

Hammad, H. M., Ashraf, M., Abbas, F., Bakhat, H. F., Qaisrani, S. A., Mubeen, M., Fahad, S., & Awais, M. (2019). Environmental factors affecting the frequency of road traffic accidents: a case study of sub-urban area of Pakistan. Environmental Science and Pollution Research, 26, 11674–11685. https://doi.org/10.1007/s11356-019-04752-8

Harris, N. L., Goldman, E., Gabris, C., Nordling, J., Minnemeyer, S., Ansari, S., Lippmann, M., Bennett, L., Raad, M., Hansen, M., & Potapov, P. (2017). Using spatial statistics to identify emerging hot spots of forest loss. Environmental Research Letters, 12(2), 024012. https://doi.org/10.1088/1748-9326/aa5a2f

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). https://doi.org/10.3390/ijgi11040260

Hissan, R. ul, Waseem, L. A., Khurshid, M., & Maqbool, M. S. (2023). The Impact of Urbanization on the Flow of Traffic in Faisalabad City Pakistan. Pakistan Journal of Humanities and Social Sciences, 11(2). https://doi.org/10.52131/pjhss.2023.1102.0432

Hussain, M., & Shi, J. (2022). Modelling and examining the influence of predictor variables on the road crashes in functionally classified vehicles in Pakistan. International Journal of Crashworthiness, 27(4), 1118–1127. https://doi.org/10.1080/13588265.2021.1909839

Ijaz, M., lan, L., Zahid, M., & Jamal, A. (2021a). A comparative study of machine learning classifiers for injury severity prediction of crashes involving three-wheeled motorized rickshaw. Accident Analysis and Prevention, 154. https://doi.org/10.1016/j.aap.2021.106094

Ijaz, M., lan, L., Zahid, M., & Jamal, A. (2021b). A comparative study of machine learning classifiers for injury severity prediction of crashes involving three-wheeled motorized rickshaw. Accident Analysis and Prevention, 154. https://doi.org/10.1016/j.aap.2021.106094

Imran, M., & Nasir, J. A. (2015). Road Traffic Accidents; Prediction in Pakistan. Professional Med J, 22(6), 705–709. www.theprofesional.com

Jacinto, G., Santos, D., Nogueira, P., Afonso, A., Silva, M., Nogueira, V., Rego, L., Quaresma, P., Saias, J., Góis, P., & Manuel, P. R. (2023). Prediction of Road Traffic Accidents on a Road in Portugal: A Mixed Artificial Intelligence, Statistics, and

Geographic Information Systems Approach. Statistics, and Geographic Information Systems Approach. https://ssrn.com/abstract=4334671

Jadaan, K., Al-Braizat, E., Al-Rafayah, S., Gammoh, H., & Abukahlil, Y. (2018). Traffic Safety in Developed and Developing Countries: A Comparative Analysis. Journal of Traffic and Logistics Engineering, 1–5. https://doi.org/10.18178/jtle.6.1.1-5

Jaiswal, A., Manoj, M., & Tiwari, G. (2024). Exploring India's Intermediate Public Transport: A Comprehensive Overview. Transportation in Developing Economies, 10(1). https://doi.org/10.1007/s40890-024-00202-4

Jayatilleke, A. U., Poudel, K. C., Dharmaratne, S. D., Jayatilleke, A. C., & Jimba, M. (2015). Factors associated with RTCs among for-hire three-wheeler drivers in Sri Lanka: a case–control study. Injury Prevention, 21(6), 374. https://doi.org/10.1136/injuryprev-2014-041501

Jurangpathy, B. S., & Rasika, S. D. (2022). Investigation of the Structural Crashworthiness of Three-wheelers.

Kamti, M. K., & Iqbal, R. (2021). Human Factors Associated with Self-reported Aberrant Behavior-A Study of Different Indian Autorickshaw Driver Age Group. 2021 International Conference on Maintenance and Intelligent Asset Management, ICMIAM 2021. https://doi.org/10.1109/ICMIAM54662.2021.9715201

Khan, I. U., Vachal, K., Ebrahimi, S., & Wadhwa, S. S. (2023). Hotspot analysis of single-vehicle lane departure crashes in North Dakota. IATSS Research, 47(1), 25–34. https://doi.org/10.1016/j.iatssr.2022.12.003

Khattak, A., Almujibah, H., Elamary, A., & Matara, C. M. (2022). Interpretable Dynamic Ensemble Selection Approach for the Prediction of Road Traffic Injury Severity: A Case Study of Pakistan's National Highway N-5. Sustainability (Switzerland), 14(19). https://doi.org/10.3390/su141912340

Kumar, P., Kumar, A., & Kumar, V. (2023). Effects of Road Conditions, Traffic Patterns, and Human-Made Features on Road Safety: Review Article. World of Science: Journal on Modern Research Methodologies, 2, 2835–3072. http://creativecommons.org/licenses/by/4.0/

Kwak, H. C., & Kho, S. (2016). Predicting crash risk and identifying crash precursors on Korean expressways using loop detector data. Accident Analysis and Prevention, 88, 9–19. https://doi.org/10.1016/j.aap.2015.12.004

La Torre, F., Meocci, M., Domenichini, L., Branzi, V., Tanzi, N., & Paliotto, A. (2019). Development of an accident prediction model for Italian freeways. Accident Analysis and Prevention, 124, 1–11. https://doi.org/10.1016/j.aap.2018.12.023

Ma, Z., Mei, G., & Cuomo, S. (2021). An analytic framework using deep learning for prediction of traffic accident injury severity based on contributing factors. Accident Analysis and Prevention, 160. https://doi.org/10.1016/j.aap.2021.106322

Mahmood, K., Batool, R., Javed, R. B., & Faizi, F. (2022). Space time dynamics of road crashes occurrence in Punjab, Pakistan: A GIS Perspective. Pakistan Journal of Scientific & Industrial Research Series A: Physical Sciences, 65(3), 260–270.

Mann, H. B. (1945). Nonparametric tests against trend. Econometrica: Journal of the Econometric Society, 245–259.

Mansoor, U., Jamal, A., Su, J., Sze, N. N., & Chen, A. (2023). Investigating the risk factors of motorcycle crash injury severity in Pakistan: Insights and policy recommendations. Transport Policy, 139, 21–38. https://doi.org/10.1016/j.tranpol.2023.05.013

Meena, S., Barwar, N., Rastogi, D., & Sharma, V. (2014). Injuries associated with cycle rickshaws accidents. Journal of Emergencies, Trauma, and Shock, 7(2), 73. https://doi.org/10.4103/0974-2700.130874

Milenković, M., & Kekic, D. (2016). Using GIS in Emergency Management. Proceedings of the International Scientific Conference - Sinteza 2016, 202–207. https://doi.org/10.15308/Sinteza-2016-202-207

Mirzaei, R., Hafezi-Nejad, N., Sadegh Sabagh, M., Ansari Moghaddam, A., Eslami, V., Rakhshani, F., & Rahimi-Movaghar, V. (2014). Dominant role of drivers' attitude in prevention of road traffic crashes: A study on knowledge, attitude, and practice of drivers in Iran. Accident Analysis and Prevention, 66, 36–42. https://doi.org/10.1016/j.aap.2014.01.013

Mohammed, S., Alkhereibi, A. H., Abulibdeh, A., Jawarneh, R. N., & Balakrishnan, P. (2023a). GIS-based spatiotemporal analysis for road traffic crashes; in support of sustainable transportation Planning. Transportation Research Interdisciplinary Perspectives, 20. https://doi.org/10.1016/j.trip.2023.100836

Mohammed, S., Alkhereibi, A. H., Abulibdeh, A., Jawarneh, R. N., & Balakrishnan, P. (2023b). GIS-based spatiotemporal analysis for road traffic crashes; in support of sustainable transportation Planning. Transportation Research Interdisciplinary Perspectives, 20. https://doi.org/10.1016/j.trip.2023.100836

Navid Tahir, M., Haworth, N., King, M., & Washington, S. (2015). Observations of Road Safety Behaviours and Practices of Motorcycle Rickshaw Drivers in Lahore, Pakistan. Proceedings of the 2015 Australasian Road Safety Conference (ARSC2015), 1–10.

Nishantha Vadysinghe, A., Herath, B., Kasun, M., Katugaha, D., Piyarathna, C., Colombage, S. M., & Vadysinghe, A. N. (2018). Injury Patterns and Causes of Death Among Occupants of Three-wheelers Succumbed to Their Injuries from Road Traffic Accidents in Sri Lanka. International Journal of Medical Toxicology and Forensic Medicine, 8(2), 55–64. https://doi.org/10.22037/ijmtfm.v0i0.20803

Ord, J. K., & Getis, A. (1995). Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. Geographical Analysis, 27(4), 286–306. https://doi.org/10.1111/j.1538-4632.1995.tb00912.x

Ospina-Mateus, H., Quintana Jiménez, L. A., Lopez-Valdes, F. J., Berrio Garcia, S., Barrero, L. H., & Sana, S. S. (2021). Extraction of decision rules using genetic algorithms and simulated annealing for prediction of severity of traffic accidents by motorcyclists. Journal of Ambient Intelligence and Humanized Computing, 12(11), 10051–10072. https://doi.org/10.1007/s12652-020-02759-5

Pakistan Automotive Manufacturers Association. (2023). Production and Sales of Vehicles.

Pervez, A., Jamal, A., & Haider Khan, S. (2024). Analyzing injury severity of three-wheeler motorized rickshaws: A correlated random parameters approach with heterogeneity in means. Accident Analysis and Prevention, 204. https://doi.org/10.1016/j.aap.2024.107651

Phun, V. K., Kato, H., & Yai, T. (2018). Traffic risk perception and behavioral intentions of paratransit users in Phnom Penh. Transportation Research Part F: Traffic Psychology and Behaviour, 55, 175–187. https://doi.org/10.1016/j.trf.2018.03.008

Plug, C., Xia, J., & Caulfield, C. (2011). Spatial and temporal visualisation techniques for crash analysis. Accident Analysis and Prevention, 43(6), 1937–1946. https://doi.org/10.1016/j.aap.2011.05.007

Pourroostaei Ardakani, S., Liang, X., Mengistu, K. T., So, R. S., Wei, X., He, B., & Cheshmehzangi, A. (2023). Road Car Accident Prediction Using a Machine-Learning-Enabled Data Analysis. Sustainability (Switzerland), 15(7). https://doi.org/10.3390/su15075939

Priye, S., & Manoj, M. (2020). Exploring usage patterns and safety perceptions of the users of electric three-wheeled paratransit in Patna, India. Case Studies on Transport Policy, 8(1), 39–48. https://doi.org/10.1016/j.cstp.2020.01.001

Rezaie Moghaddam, F., Afandizadeh, S., & Ziyadi, M. (2011). Prediction of accident severity using artificial neural networks. Int J Civ Eng, 9(1), 41–48.

Rezapour, M., Nazneen, S., & Ksaibati, K. (2020). Application of deep learning techniques in predicting motorcycle crash severity. Engineering Reports, 2(7). https://doi.org/10.1002/eng2.12175

Rolison, J. J., Regev, S., Moutari, S., & Feeney, A. (2018). What are the factors that contribute to road accidents? An assessment of law enforcement views, ordinary drivers' opinions, and road accident records. Accident Analysis and Prevention, 115, 11–24. https://doi.org/10.1016/j.aap.2018.02.025

Saha, B., Fatmi, M. R., & Rahman, M. M. (2021). Traffic crashes in Dhaka, Bangladesh: analysing crashes involving unconventional modes, pedestrians and public transit. International Journal of Injury Control and Safety Promotion, 28(3), 347–359. https://doi.org/10.1080/17457300.2021.1928230

Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). Sensitivity analysis in practice: a guide to assessing scientific models (Vol. 1). Wiley Online Library.

Sameen, M. I., & Pradhan, B. (2017). Severity prediction of traffic accidents with recurrent neural networks. Applied Sciences (Switzerland), 7(6). https://doi.org/10.3390/app7060476

Schmucker, U., Dandona, R., Kumar, G. A., & Dandona, L. (2011). Crashes involving motorised rickshaws in urban India: Characteristics and injury patterns. Injury, 42(1), 104–111. https://doi.org/10.1016/j.injury.2009.10.049

Sindha, J., Chakraborty, B., & Chakravarty, D. (2015). Rigid body modeling of three wheel vehicle to determine the dynamic stability-A practical approach. 2015 IEEE International Transportation Electrification Conference (ITEC), 1–8.

Singh, R. K., & Suman, R. K. (2012). Accident Analysis and Prediction of Model on National Highways. International Journal of Advanced Technology in Civil Engineering, 1(2), 25–30.

Somasundaraswaran, A. K., & Tay, R. (2006). Self-reported differences between crash-involved and non-crash-involved three-wheeler drivers in Sri Lanka. IATSS RESEARCH, 30(1), 68–72.

Srikanth, K. M., & Prakash, R. V. (2009). Assessing the structural crashworthiness of a three-wheeler passenger vehicle. ICORD 09: Proceedings of the 2nd International Conference on Research into Design, Bangalore, India 07.-09.01.2009, 152–159.

Starkey, P., Batool, Z., Younis, E. M. W., Rehman, A. U., & Ali, M. S. (2021). Motorcycle three-wheelers in Pakistan: Low-cost rural transport services, crucial for women's mobility. Transportation Research Interdisciplinary Perspectives, 12. https://doi.org/10.1016/j.trip.2021.100479

Starkey, P., Batool, Z., & Younis, M. W. (2019). The Expansion of Three-Wheeler Transport Services: The Case of Qingqis in Pakistan. A Paper Presented at the 26th World Road Congress.

Subhan, F., Zhou, H., Zhao, S., Naeem, M. M., & Sulaiman, M. (2021). Modeling non-fatal road crash injuries for Pakistan using aggregate data. Advances in Transportation Studies, 53, 147–166.

Tamakloe, R. (2023). Risk Factors Influencing Fatal Powered Two-Wheeler At-Fault and Not-at-Fault Crashes: An Application of Spatio-Temporal Hotspot and Association

Rule Mining Techniques. Informatics, 10(2). https://doi.org/10.3390/informatics10020043

Ullah, H., Farooq, A., & Shah, A. A. (2021). An Empirical Assessment of Factors Influencing Injury Severities of Motor Vehicle Crashes on National Highways of Pakistan. Journal of Advanced Transportation, 2021. https://doi.org/10.1155/2021/6358321

Umam Syaliman, K. (2021). Enhance the accuracy of k-Nearest Neighbor (k-NN) for unbalanced class data using Synthetic Minority Oversampling Technique (SMOTE) and Gain Ratio (GR). JURNAL INFOKUM, 10(1). http://infor.seaninstitute.org/index.php/infokum/index

Wang, J., Liu, B., Fu, T., Liu, S., & Stipancic, J. (2019). Modeling when and where a secondary accident occurs. Accident Analysis and Prevention, 130, 160–166. https://doi.org/10.1016/j.aap.2018.01.024

World Health Organization. (2015). Global status report on road safety.

World Health Organization. (2017). Powered two-and three-wheeler safety a road safety manual for decision-makers and practitioners.

World Health Organization. (2018). Global status report on road safety.

Wu, P., Chen, T., Wong, Y. D., Meng, X., Wang, X., & Liu, W. (2023). Exploring key spatio-temporal features of crash risk hot spots on urban road network: A machine learning approach. Transportation Research Part A: Policy and Practice, 173. https://doi.org/10.1016/j.tra.2023.103717

APPENDIX A: SYNAPTIC WEIGHTS BY NEURAL NETWORK

Parameter Estimates

Predictor		Predicted									
		Hidden Layer 1									
		H(1:1)	H(1:2)	H(1:3)	H(1:4)	H(1:5)	H(1:6)	H(1:7)	H(1:8)	[Injury Id=1]	[Injury Id=2]
Input Layer	(Bias)	0.251	-0.148	1.373	-0.048	-0.073	-0.065	-0.52	0.23		
	Responsetime	0.184	0.536	0.329	-0.736	0.467	-0.434	-0.297	0.266		
	Avg_temp	0.477	0.89	-1.314	0.938	0.172	0.367	-0.367	0.203		
	NumberofVictims	-0.169	-0.558	-0.097	1.204	0.02	0.45	0.018	-0.203		
	Day	0.392	0.864	0.524	-0.646	-0.536	0.403	-0.381	-0.015		
	Season	-0.474	0.55	0.92	0.065	0.473	-0.783	-0.235	0.307		
	Time	-0.156	-0.035	0.082	-0.041	-0.851	0.058	0.4	-0.313		
	Peakhours	0.41	0.011	0.555	0.284	-0.586	0.58	0.505	-0.645		
	Gender	0.591	-0.517	0.441	0.404	-0.084	2.06	-0.424	0.062		
	Age	-0.571	0.916	0.144	-0.186	0.09	0.159	-0.489	-0.422		
	Edu	0.159	-0.656	0.488	0.203	-1.133	-0.395	-0.282	-0.035		
	RTAcauseid	0.026	0.874	-0.386	-0.297	-0.501	-0.124	0.734	-0.458		
	VictimType	-0.632	-0.272	-0.323	-0.459	-0.542	2.103	0.248	-0.551		
	TotalVehInvolved	-0.427	-0.637	0.204	0.089	0.331	0.23	0.173	0.544		
	Weather	-0.33	0.147	0.071	0.372	0.645	-0.639	1.019	0.113		
Hidden Layer 1	(Bias)									0.132	-0.127
	H(1:1)									0.518	-0.548
	H(1:2)									-1.018	0.888
	H(1:3)									-1.182	1.112
	H(1:4)									-0.799	0.773
	H(1:5)									0.527	-0.454
	H(1:6)									-1.01	0.952
	H(1:7)									-0.515	0.456
	H(1:8)									-0.283	0.191