

FRAMEWORK FOR ANALYSIS OF HIT-AND-RUN ROAD TRAFFIC CRASHES

ALOK NIKHIL JHA



**TRANSPORTATION RESEARCH AND INJURY
PREVENTION CENTER**

INDIAN INSTITUTE OF TECHNOLOGY DELHI

AUGUST 2023

© Indian Institute of Technology Delhi (IITD), New Delhi, 2023

FRAMEWORK FOR ANALYSIS OF HIT-AND-RUN ROAD TRAFFIC CRASHES

by

ALOK NIKHIL JHA

TRANSPORTATION RESEARCH AND INJURY PREVENTION CENTER

Submitted

In fulfilment of the requirements of the degree of Doctor of Philosophy

to the



Indian Institute of Technology Delhi

August 2023

CERTIFICATE

This is to certify that the thesis titled “Framework for Analysis of Hit-and-run Road Traffic Crashes” being submitted by Alok Nikhil Jha to the Indian Institute of Technology Delhi for the award of the degree of Doctor of Philosophy is a record of the original bonafide research work carried out by him under our guidance. The thesis work, in our opinion, has reached the requisite standard fulfilling the requirements for the degree of Doctor of Philosophy. The results contained in the thesis have not been submitted in part or full to any other University or Institute for the award of any degree or diploma.



Dr. Niladri Chatterjee
Professor
Department of Mathematics
Indian Institute of Technology Delhi
New Delhi – 110016



Dr. Geetam Tiwari
Professor
Transport Research (TRIP) Center
Indian Institute of Technology Delhi
New Delhi – 110016

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my Research Supervisors, Prof. Niladri Chatterjee (Soumitra Dutta Chair Professor in Artificial Intelligence and Professor, Dept. of Mathematics, School of IT, School of AI at IIT Delhi) and Prof. Geetam Tiwari (Professor & Head, TRIP Center, at IIT Delhi), for providing me with the invaluable opportunity to pursue this research work under their guidance. Their belief in me and the freedom they offered me to find my path were crucial to my success as an independent researcher. I am grateful for their guidance and methodology that helped me throughout this journey. Additionally, I would like to thank Sh. Mahesh Ji for his administrative support.

I want to thank Prof. Aparna Mehra, who guided me toward going for a PhD to channel my research ideas. I am indebted to my friends and colleagues who made this journey easier and more enjoyable. My three musketeers, Abhijeet Parmar, Ajay Kumar, and Sachin Kumar, provided invaluable insights and criticisms that helped me approach the problem more clearly. I am also grateful to my classmates and friends who supported and motivated me in various ways, including Varun, Kulbhushan, Dhaka, Richa, Aditya, Krishna, Priyanka, and Swati. I am thankful to Neelam for her encouragement and to Pooja for suggestions and inputs.

I am deeply indebted to my family, especially my parents, for their unwavering support and encouragement. I want to thank my mother Smt. Poonam Jha and father Sh. Kanti Dhar Jha and my sisters Archana, Vandana, and Usha, for their continuous motivation. My brother Nitish and brother-in-law Sh. Vishnu Kant Jha also deserves my gratitude for their discussions on various research trends. Additionally, I am thankful to my brother-in-law Dr. Pravin Choudhary, who provided valuable guidance and input. I am also thankful to IIIT Delhi for its support. Finally, I dedicate my research to my nephew, Chandrachur, who died in a hit-and-run crash at a young age.

I could not have completed this work without the support and help of all the individuals mentioned above, as well as the blessings of the almighty.

Alok Nikhil Jha

ABSTRACT

Transportation is crucial for the development of any country, and road transportation plays a significant role among all the modes of transportation. One of the most common challenges in transportation is safety, and road safety is of utmost importance due to the high number of fatalities, injuries, and property damage resulting from road crashes. The UN General Assembly Decade of Action for Road Safety 2021-2030 aims to prevent at least 50% of road traffic deaths and injuries by 2030, and hit-and-run crashes pose a significant challenge to achieving this goal. Hit-and-run crashes are dangerous because the responsible vehicles leave the crash location without reporting the incident or helping the victims. This significantly increases the risks of severe injuries and fatalities due to emergency assistance delays. Despite hit-and-run being designated as criminal offence, these crashes are common, and the analysis of such incidents is challenging due to the unavailability of information about the impacting vehicles and difficult to analyze and address this.

This thesis presents a framework to predict hit-and-run crash-causing vehicles and fills the research gap about incomplete crash database due to unknown vehicles of hit-and-run crashes. Identifying unknown vehicles involved in hit-and-run crashes improves the quality of road crash data, and crucial timings and potential impacting vehicles and victims' portfolios are also studied. The framework is based on machine learning techniques, and its results are validated and verified through temporal trends, observations, and statistics. The results show that the possible hit-and-run vehicles predicted through machine learning models are synchronous to observations during that period. Other findings on vulnerable timings and days of fatal hit-and-run crashes are similar to some established research in the area. The methodology is highly scalable and capable of completing a crash database, thereby identifying useful information previously ignored. The thesis explores the impact of hit-and-run crashes and proposes mechanisms to minimize them, which can help achieve the UN's ambitious targets for road safety. The proposed framework can serve as a basis for future research and road safety planning, highlighting the importance of addressing hit-and-run crashes to improve road safety.

सारांश

किसी भी देश के विकास के लिए परिवहन महत्वपूर्ण है और परिवहन के सभी साधनों में सड़क परिवहन एक महत्वपूर्ण भूमिका निभाता है। परिवहन में सबसे आम चुनौतियों में से एक सुरक्षा है, और सड़क दुर्घटनाओं के परिणामस्वरूप होने वाली मौतों, चोटों और संपत्ति की क्षति की उच्च संख्या के कारण सड़क सुरक्षा अत्यंत महत्वपूर्ण है। सड़क सुरक्षा के लिए संयुक्त राष्ट्र महासभा की कार्रवाई दशक 2021-2030 का लक्ष्य 2030 तक कम से कम 50% सड़क यातायात मौतों और चोटों को रोकना है, और हिट-एंड-रन दुर्घटनाएं इस लक्ष्य को प्राप्त करने के लिए एक महत्वपूर्ण चुनौती पेश करती हैं। हिट-एंड-रन दुर्घटनाएं खतरनाक होती हैं क्योंकि जिम्मेदार वाहन घटना की रिपोर्ट किए बिना या पीड़ितों की मदद किए बिना दुर्घटना स्थल छोड़ देते हैं। इससे आपातकालीन सहायता में देरी के कारण गंभीर चोटों और मृत्यु का जोखिम काफी बढ़ जाता है। हिट-एंड-रन को आपराधिक अपराध के रूप में नामित किए जाने के बावजूद, ये दुर्घटनाएं आम हैं, और प्रभावित करने वाले वाहनों के बारे में जानकारी की अनुपलब्धता और इसका विश्लेषण करना और इसका समाधान करना मुश्किल होने के कारण ऐसी घटनाओं का विश्लेषण चुनौतीपूर्ण है।

यह थीसिस हिट-एंड-रन दुर्घटना का कारण बनने वाले वाहनों की भविष्यवाणी करने के लिए एक रूपरेखा प्रस्तुत करती है और हिट-एंड-रन दुर्घटनाओं के अज्ञात वाहनों के कारण अपूर्ण दुर्घटना डेटाबेस के बारे में शोध अंतर को भरती है। हिट-एंड-रन दुर्घटनाओं में शामिल अज्ञात वाहनों की पहचान करने से सड़क दुर्घटना डेटा की गुणवत्ता में सुधार होता है, और महत्वपूर्ण समय और संभावित प्रभावित करने वाले वाहनों और पीड़ितों के पोर्टफोलियो का भी अध्ययन किया जाता है। रूपरेखा मशीन लर्निंग तकनीकों पर आधारित है, और इसके परिणामों को अस्थायी रुझानों, टिप्पणियों और आंकड़ों के माध्यम से मान्य और सत्यापित किया जाता है। परिणाम बताते हैं कि मशीन लर्निंग मॉडल के माध्यम से संभावित हिट-एंड-रन वाहनों की भविष्यवाणी उस अवधि के दौरान टिप्पणियों के अनुरूप है। घातक हिट-एंड-रन दुर्घटनाओं के संवेदनशील समय और दिनों पर अन्य निष्कर्ष क्षेत्र में कुछ स्थापित शोध के समान हैं। कार्यप्रणाली अत्यधिक स्केलेबल है और क्रैश डेटाबेस को पूरा करने में सक्षम है, जिससे पहले नजरअंदाज की गई उपयोगी जानकारी की पहचान की जा सकती है। थीसिस हिट-एंड-रन दुर्घटनाओं के प्रभाव की पड़ताल करती है और उन्हें कम करने के लिए तंत्र का प्रस्ताव करती है, जो सड़क सुरक्षा के लिए संयुक्त राष्ट्र के महत्वाकांक्षी लक्ष्यों को प्राप्त करने में मदद कर सकती है। प्रस्तावित रूपरेखा भविष्य के अनुसंधान और सड़क सुरक्षा योजना के लिए आधार के रूप में काम कर सकती है, जो सड़क सुरक्षा में सुधार के लिए हिट-एंड-रन दुर्घटनाओं को संबोधित करने के महत्व पर प्रकाश डालती है।

TABLE OF CONTENTS

CERTIFICATE	i
ACKNOWLEDGEMENTS	ii
ABSTRACT.....	iii
LIST OF FIGURES	viii
LIST OF TABLES	xi
CHAPTER 1: INTRODUCTION	1
1.1. Road Safety and Traffic Crashes	1
1.2. Hit-and-Run Crashes	2
1.3. Motivation and Relevance of Research.....	4
1.4. Research Objective	5
1.5. Methodology of Research.....	5
1.6. Structure of Thesis	7
CHAPTER 2: LITERATURE REVIEW	9
2.1. Introduction.....	9
2.2. Background.....	10
2.3. Hit-and-run Crashes.....	14
2.4. Past Efforts in Reducing Hit-and-Run Road Crashes	16
2.5. Gap Analysis	27
2.6. Conclusion	29
CHAPTER 3: ROAD TRAFFIC CRASH DATA	31
3.1. Road Traffic Crash Data.....	31
3.2. Recording of Crash in the USA.....	33
3.2.1. General Estimate System or GES	34
3.2.2. Crashworthiness Data System or CDS	34
3.2.3. Crash Report Sampling System (CRSS).....	36
3.2.4. Crash Investigation Sampling System (CISS)	37
3.2.5. Fatality and Crash Reporting System (FARS).....	37
3.3. Recording of Crash in European Union	39

3.3.1. Common Crash Dataset (CAdaS)	40
3.4. Recording of Crash in India	42
3.5. Conclusion	46
CHAPTER 4: MISSING DATA IN ROAD TRAFFIC CRASHES	48
4.1. Introduction.....	48
4.2. Missing data and Errors	49
4.3. Missing Data: Interpretation and Analysis	51
4.4. Data Collection Study	55
4.5. Findings and Analysis	57
4.5.1. Crash-Related Data	57
4.5.2. Vehicle-Related Data	60
4.5.3. Victim Related Data	63
4.5.4. Analysis.....	67
4.6. Conclusions and Future Work	68
4.7. Contributions.....	69
CHAPTER 5: IDENTIFICATION OF OFFENDERS IN HIT-AND-RUN CRASHES	71
5.1. Introduction.....	71
5.2. Methodology	72
5.3. Machine Learning Models and Relevance.....	73
5.3.1. Classification algorithms	73
5.3.2. Relevance of the Machine Learning Approach	88
5.3.3. Finding Suitable Models & Cross Validation	90
5.4. Road Crash Data Description and Details	93
5.5. Data Insights & Dependency	94
5.5.1. Crash Data Patterns.....	94
5.5.2. Crash period and vehicle type dependency	97
5.6. Results and Discussions.....	98
5.6.1. Results.....	98
5.6.2. Discussions & Analysis	101
5.7. Conclusion	106
5.8. Contributions.....	107
CHAPTER 6: EFFECT OF HIT-AND-RUN CRASHES ON CRASH DATA	109

6.1.	Introduction.....	109
6.2.	Data Description and Insights.....	109
6.2.1.	Data Description.....	109
6.2.2.	Insights of Hit-and-Run Road Crashes in Four Cities.....	111
6.2.2.1.	Hyderabad.....	111
6.2.2.2.	Shimla.....	113
6.2.2.3.	Dehradun.....	114
6.2.2.4.	Chandigarh.....	116
6.2.2.5.	Summary of crash data from four cities.....	117
6.3.	Methodology and Approach	118
6.4.	Results and Discussions.....	119
6.4.1.	Results.....	119
6.4.1.1.	Findings in Hyderabad.....	119
6.4.1.2.	Findings for Shimla	122
6.4.1.3.	Findings for Dehradun	125
6.4.1.4.	Findings for Chandigarh.....	128
6.4.2.	Analysis.....	131
6.5.	Conclusions.....	133
6.6.	Contribution.....	133
CHAPTER 7: INSIGHTS OF FATAL HIT-AND-RUN CRASHES.....		134
7.1.	Introduction.....	134
7.2.	Methodology	135
7.3.	Data Collection and Analysis	136
7.3.1.	Data Collection and Cleaning.....	136
7.3.2.	Data Schema.....	137
7.3.3.	Assumptions	138
7.3.4.	Computation	139
7.4.	Results.....	139
7.4.1.	Identification of Best Model.....	139
7.4.2.	Unknown vehicles causing a hit-and-run.....	140
7.5.	Analysis and Discussions	141
7.5.1.	Findings from hit-and-run crashes (Phase – 2).....	141

7.5.2. Impact and Crash Patterns Analysis (Phase – 3).....	144
7.6. Conclusion	155
7.7. Contributions.....	156
CHAPTER 8: CONCLUSION AND FUTURE WORK	157
8.1. Concluding Remarks	157
8.2. Contributions.....	158
8.3. Recommendations	160
8.4. Limitations	161
8.5. Future work.....	163
References	165
Bio-Data.....	176
Publications.....	177

LIST OF FIGURES

Figure 1.1: Approach followed in the study-----	6
Figure 1.2: : Evaluation and comparison of machine learning models -----	6
Figure 2.1: Road crash as a leading cause of death -----	11
Figure 2.2: Hit-and-run Crashes reported in India -----	15
Figure 3.1: Crash reporting workflow in India -----	42
Figure 4.1: Type of Errors in crash data -----	50
Figure 5.1: Proposed approach -----	73
Figure 5.2: Structure of decision tree -----	79
Figure 5.3: Classification using support vector machine-----	82
Figure 5.4: Structure of random trees -----	84
Figure 5.5: Layers in neural network-----	85
Figure 5.6: Representation of perceptron -----	86
Figure 5.7: Illustration of K-Fold process -----	92
Figure 5.8: Actual crash data and cleaned crash data -----	95
Figure 5.9: Actual versus cleaned crash records for 3 periods -----	95
Figure 5.10: Impacting vehicles in period 1-----	96
Figure 5.11: Impacting vehicles in period 2 -----	97
Figure 5.12: Impacting vehicles in period 3 -----	97
Figure 5.13: Hit-and-run vehicles in Period 1 -----	100
Figure 5.14: Hit-and-run vehicles in Period 2 -----	101
Figure 5.15: Hit-and-run vehicles in Period 3 -----	101
Figure 5.16: Hit-and-run vehicles in for all periods -----	101

Figure 5.17: Year-wise summary of the hit-and-run crash-causing impacting vehicles -----	106
Figure 6.1: Impacting vehicles composition in Hyderabad -----	112
Figure 6.2: Impacted road users in Hyderabad -----	112
Figure 6.3: Impacting vehicles in hit-and-run crashes in Hyderabad-----	112
Figure 6.4: Victims of hit-and-run crashes in Hyderabad-----	112
Figure 6.5: Crash-causing vehicles in Shimla -----	114
Figure 6.6: Victims of road crashes in Shimla-----	114
Figure 6.7: Impacting vehicles in hit-and-run crash in Shimla-----	114
Figure 6.8: Victims of hit-and-run crashes in Shimla -----	114
Figure 6.9: Crash-causing vehicles in Dehradun-----	115
Figure 6.10: Victims of road crashes in Dehradun-----	115
Figure 6.11: Impacting vehicles in hit-and-run crashes-----	115
Figure 6.12: Victims of hit-and-run crashes -----	115
Figure 6.13: Crash-causing vehicles in Chandigarh-----	116
Figure 6.14: Victims of road crashes in Chandigarh-----	116
Figure 6.15: Impacting vehicles in hit-and-run crashes in Chandigarh-----	117
Figure 6.16: Victims of hit-and-run crashes in Chandigarh-----	117
Figure 6.17: Proposed Model-----	118
Figure 6.18: Impacting vehicles in Hyderabad (with unknown) -----	120
Figure 6.19: Impacting vehicles in Hyderabad (unknown predicted) -----	120
Figure 6.20: Hit-and-run impacting vehicles in Hyderabad (with unknown) -----	121
Figure 6.21: Hit-and-run impacting vehicles in Hyderabad (unknown predicted) -----	121
Figure 6.22: Impacting vehicles in Shimla (with unknown) -----	123
Figure 6.23: Impacting vehicles in Shimla (unknown predicted) -----	123

Figure 6.24: Hit-and-run impacting vehicles in Shimla (with unknown) -----	124
Figure 6.25: Hit-and-run impacting vehicles in Shimla (unknown predicted) -----	124
Figure 6.26: Impacting vehicles in Dehradun (with unknown) -----	126
Figure 6.27: Impacting vehicles in Dehradun (unknown predicted) -----	126
Figure 6.28: Hit-and-run impacting vehicles in Dehradun (with unknown) -----	127
Figure 6.29: Hit-and-run impacting vehicles in Dehradun (unknown predicted) -----	127
Figure 6.30: Impacting vehicles in Chandigarh (with unknown) -----	129
Figure 6.31: Impacting vehicles in Chandigarh (unknown predicted) -----	129
Figure 6.32: Hit-and-run impacting vehicles in Chandigarh (with unknown) -----	130
Figure 6.33: Hit-and-run impacting vehicles in Chandigarh (unknown predicted) -----	130
Figure 7.1: Framework for hit-and-run crashes -----	135
Figure 7.2: Workflow for computation of unknown vehicles -----	139
Figure 7.3: Vehicle share having a high possibility of causing a hit-and-run accident -----	146
Figure 7.4: City-wise summary of victims of Hit-and-run accidents -----	148
Figure 7.5: Hit-and-run share in different timings-----	149
Figure 7.6: Vulnerable timing for hit-and-run accidents-----	150
Figure 7.7: Hit-and-run crashes on different days-----	151
Figure 7.8: Vulnerable days for hit-and-run-----	152

LIST OF TABLES

Table 2.1: Data, methods, and key factors in previous studies -----	22
Table 3.1: Vehicle-related data in FARS-----	38
Table 3.2: Crash level data in FARS -----	39
Table 3.3: CADaS Variables & Values -----	40
Table 3.4: Crash-related variables -----	41
Table 3.5: Road variables -----	41
Table 3.6: Traffic variables -----	41
Table 3.7: Person variables -----	41
Table 3.8: Crash variables in India -----	44
Table 4.1: Missing data patterns -----	51
Table 4.2: Missing data type -----	52
Table 4.3: Traffic fatality study -----	55
Table 4.4: Crash records in Indian Cities -----	56
Table 4.5: Proportion of missing data for crash variables -----	57
Table 4.6: Missing type from recorded data for crash-related variables -----	60
Table 4.7: Proportion of missing data in vehicle-related attributes -----	61
Table 4.8: Missing type from recorded data for Vehicle related variables -----	63
Table 4.9: Proportion of missing data in Pedestrian and NMT victims -----	64
Table 4.10: Missing type from recorded data for Pedestrian Victims related variables -----	65
Table 4.11: Proportion of missing data in Motorized Victims -----	66
Table 4.12: Missing type from recorded data for Victims related variables -----	66
Table 4.13: Average missing data in accident records -----	68
Table 5.1: Variables in Road Crash Data from Delhi -----	94

Table 5.2: Impacting vehicles for 3 Periods -----	96
Table 5.3: Performance of each model for periods 1, 2 and 3 -----	99
Table 5.4: Impacting vehicles causing hit-and-run crashes -----	100
Table 5.5: Contingency matrix generated -----	103
Table 5.6: Testing null hypothesis -----	104
Table 6.1: Variables in the crash data-----	110
Table 6.2: Summary of crashes data for the four cities-----	117
Table 6.3: Performance of classification models for prediction of unknown vehicles-----	119
Table 6.4: Actual vs modified profile of impacting vehicles in Hyderabad -----	122
Table 6.5: Actual vs modified profile of impacting vehicles in Shimla -----	125
Table 6.6: Actual vs modified profile of impacting vehicles in Dehradun -----	128
Table 6.7: Actual vs modified profile of impacting vehicles in Chandigarh -----	131
Table 7.1: Road crash data from nine cities-----	137
Table 7.2: Variables in the crash data-----	138
Table 7.3: Performance of various models-----	140
Table 7.4: Predicted unknown vehicles and fatalities for all cities-----	141
Table 7.5: Summary of hit-and-run impacting vehicles and victims with count -----	142
Table 7.6: Summary of hit-and-run crashes timings with respective count-----	143
Table 7.7: Unknown vehicles predicted-----	145
Table 7.8: Impact on hit-and-run crash composition before/after predicting unknown ----	146
Table 7.9: Impact on portfolio on Hit-and-Run fatal crashes on various times -----	149
Table 7.10: Impact on portfolio on Hit-and-Run fatal crashes on various days-----	151
Table 7.11: Impact indices for computation of impact of hit-and-run-----	154
Table 7.12: Impact of hit-and-run on cities-----	154