APPLICATION OF DATA MINING IN ROAD TRAFFIC ACCIDENT ANALYSIS

by

Md. Asif Raihan

MASTER OF SCIENCE IN CIVIL ENGINEERING (TRANSPORTATION)

Department of Civil Engineering

BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY

June 2013

M. Sc.
Engg.

(Civil & Transportation)

M. A. Raihan

June
2013

Application of Data Mining in Road Traffic Accident Analysis

by

Md. Asif Raihan

MASTER OF SCIENCE IN CIVIL ENGINEERING
(TRANSPORTATION)

Department of Civil Engineering

BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY

June 2013

The thesis titled 'Application of Data Mining in Road Traffic Accident Analysis' submitted by Md. Asif Raihan, Roll No: 040804045 (P), Session: April 2008 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Master of Science in Civil Engineering (Transportation) on June 29, 2013.

BOARD OF EXAMINERS

Dr. Tanweer Hasan Professor, Department of Civil Engineering Director, Accident Research Institute (ARI) Bangladesh University of Engineering and Technology (BUET) Dhaka 1000, Bangladesh	Chairman
Dr. Md. Mujibur Rahman Professor and Head Department of Civil Engineering Bangladesh University of Engineering and Technology (BUET) Dhaka 1000, Bangladesh	Member (Ex-officio)
Dr. Md. Mizanur Rahman Professor Department of Civil Engineering Bangladesh University of Engineering and Technology (BUET) Dhaka 1000, Bangladesh	Member
Farzana Rahman Dr. Farzana Rahman Assistant Professor Presidency University Dhaka, Bangladesh	Member (External)

DECLARATION

This is to certify that the thesis entitled 'Application of Data Mining in Road Traffic Accident Analysis' submitted to the Department of Civil Engineering, Bangladesh University of Engineering and Technology (BUET) in partial fulfillment of the requirement for the degree of Master of Science in Civil Engineering (Transportation) is a record of original research work done by me (Md. Asif Raihan) under the supervision of Dr. Tanweer Hasan, Professor, Department of Civil Engineering, BUET and Director, Accident Research Institute (ARI), BUET, and the thesis has not been submitted elsewhere for any award/degree/diploma/fellowship or for any other purpose.

Asif Raihan

Md. Asif Raihan
June 29, 2013

ACKNOWLEDGEMENT

First and foremost, I would like to express my sincere and humble gratitude to the Almighty Allah, the most Beneficent and the most Merciful for enabling me to carry out the research work and complete it without any hitch.

I would like to express my reverential gratitude to my supervisor Dr. Tanweer Hasan, Professor, Department of Civil Engineering, BUET and Director, Accident Research Institute (ARI), for his extraordinary, invaluable suggestions and continuous encouragement in the preparation of this thesis. Without his guidance and persistent support this research would have not been possible to complete.

My profound and heartfelt thanks to my mentor Moinul Bhai (Dr. Moinul Hossain, Adjunct Professor, Department of Chemical Engineering and Applied Chemistry, Faculty of Applied Science and Engineering, University of Toronto) for his patience, meticulous care, continuous inspiration, valuable advice and dynamic guidance all of which molded this study into a reality.

I would also like to thank the faculties, officers and office personnel of ARI for their cooperation.

And finally, I would like to thank the inseparable part of my life my parents Nilufar Fatema and Mohammad Farhad Hossain, my law-parents Dr. Zinat Mahrukh Banu and Professor Dr. Md. Anwarul Islam, my younger brother Shabib and last but not the least my wife Shahrin who endured this long process with me, always offering affectionate support, patience, careful assistance and inspiration.

ABSTRACT

The existing road traffic accident (RTA) analysis system in Bangladesh is more focused onto record management and basic data analysis i.e. characteristics analysis purposes rather than using it as a source of intelligence. Although MAAP based accident database constitute the respiratory for RTA information of the country, its application is constrained by a number of limitations. However, most of the previous studies focused on a few risk factors, some specific road users or certain types of crashes; and therefore the important factors affecting injury or crash severity have not been completely recognized yet.

Data mining (DM) has the potential to eliminate RTA data related deficiencies as well as statistical limitations. Even DM is able to quantify multiple relationships, which provides the insight for policy level decisions. Therefore, DM has been utilized in this thesis to elicit reasonable, novel, and interesting facts and also to confirm some perceived facts using RTA data (1998-2010) from ARI, BUET. Several DM algorithms have been adopted for the study. At first, hierarchical clustering (HC) methodology was employed to form natural data groups and to identify hazardous clusters; then random forest (RF) was applied to identify, rank, and thus select a subset of variables from a large variable space, to be considered for this study. Finally, classification and regression trees (CART) have been allowed to investigate the accident severity mechanism of the hazardous clusters.

Nearly 10 percent of the pedestrian accidents are triggered by other accident/collision types, which indicate that may be pedestrians are not only the victims but also a stimulating factor for some accidents. Dividers in urban areas have been found quite effective in reducing fatal (38.23% fatal vs 57.78% fatal where there are no dividers) pedestrian accidents. Traffic control systems especially police controlled traffic control system in urban areas have been identified as persuasive in reducing pedestrian fatal accidents (in some cases 0% fatal incidences). Geometric sections without police controlled traffic control system have been acknowledged as a bracing factor for fatal pedestrian accidents.

Straight and flat geometric sections of roadways have generated more double vehicle fatal accidents (more than 80% accidents are fatal) than other types (e.g. curve only, slope only, curve and slope and crest) of geometric sections (nearly 70% fatal accidents). The latter part of the previous finding got worse when the sections were associated with head on, right angle, overturn, hit object in road and hit animal type of collisions (76.22% fatal); or occurred on national and regional highways or feeder roads (71% fatal); or during dawn/dusk and night (unlit) lighting condition (90.91% fatal); or in daylight or night (lit) light condition but with no or centerline marking traffic control system (75.21% fatal).

Head on, right angle, side swipe, hit object in road, and hit object off road collision types affiliated with curve only, slope only, and curve and slope geometric sections of the roadways produced 85.29 percent fatal single vehicle crashes. Dawn/dusk and night (unlit) lighting condition attributed 87.88 percent single vehicle fatal accidents. Brick and earthen road surfaces have generated 86.67% fatal single vehicle crashes even in daylight and night (lit) condition. On the contrary, sealed surface even affiliated with rainy weather has ensued less fatal single vehicle crashes (58.82% non-fatal crashes).Wet and flooded surface conditions of roads have resulted in 94.74 percent fatal single vehicle crashes. Nevertheless, one-way routes concomitant with dry and muddy surface prompted only 20 percent fatal cases as always perceived; whereas in case of two-way roads it shoots up to 86.54 percent fatal single vehicle accidents.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ADU	Accident Data Unit
ARF	Accident Report Form
ARI	Accident Research Institute
BRTA	Bangladesh Road Transport Authority
BUET	Bangladesh University of Engineering and Technology
CART	Classification and Regression Tree
DFID	Department for International Development
DM	Data Mining
DMP	Dhaka Metropolitan Police
FIR	First Information Report
GB	Gigabyte
GDP	Gross Domestic Product
GIS	Geographical Information System
HC	Hierarchical Clustering
HQ	Headquarters
HRL	Hazardous Road Locations
IDC	Institutional Development Component
MAAP	Microcomputer Accident Analysis Package
MS	Microsoft
PC	Personal Computer
RAM	Random Access Memory
RF	Random Forest
RRMP	Road Rehabilitation and Maintenance Project
RSC	Road Safety Cell
RTA	Road Traffic Accident
RUM	Road User Movement
SQL	Structured Query Language
TRL	Transport Research Laboratory
UK	United Kingdom
WHO	World Health Organization

CHAPTER 1
INTRODUCTION

1.1 Background and Motivation

Road traffic accidents (RTAs) are a major public health concern, resulting over 1.2 million deaths and between 20 and 50 million non-fatal injuries worldwide each year. Low-income and middle-income countries have higher road traffic fatality rates (21.5 and 19.5 per 1,00,000 population, respectively) than high-income countries (10.3 per 1,00,000). Over 90 percent of the world's fatalities on roads occur in low-income and middle-income countries, which have only 48 percent of the world's registered vehicles. The global losses due to road traffic injuries are estimated to be US$ 518 billion and cost governments between 1% and 3% of their gross national product – more than the total amount that these countries receive in development assistance. While road traffic death rates in many high-income countries have stabilized or declined in recent decades, data suggest that in most regions of the world the global epidemic of traffic injuries is still increasing. It has been estimated that, unless immediate action is taken, road deaths will rise to the fifth leading cause of death by 2030, resulting in an estimated 2.4 million fatalities per year [WHO, 2009].

Bangladesh in particular experiences one of the highest rate of such accidents. According to police reported statistics around 4,000 people die through RTAs in Bangladesh each year. It is estimated that the actual fatalities could well be 10,000–12,000 each year taking consideration of underreporting and definitional inconsistencies. In economic terms, road accidents in Bangladesh are costing the community nearly 2 percent of GDP. This is, of course, a huge sum that the nation can ill afford to lose [Hoque *et al.*, 2008]. Thus, methods to reduce accident severity are of great interest to traffic agencies and to public at large.

Research based on comprehensive analysis of the causes of accidents and design of appropriate engineering solution is the key to successful endeavor. Scientific investigations and implementation of commensurate technical measures are contingent upon the availability of ample information on accident which includes data on vehicle, roadway, environment, users and victims as well. In Bangladesh police is the core organization for accident data collection and storage [Alam *et al.*, 2006]. Their accident database system is computerized through the application of Microcomputer Accident Analysis Package (MAAP) which is developed by the Transport Research Laboratory (TRL) of the United Kingdom (UK) specifically for the storage and analysis of accident data. The Accident Research Institute (ARI) of Bangladesh University of Engineering and Technology (BUET) essentially uses the MAAP database for research purposes. This database was transferred to the institute with institutional collaboration of the Road Safety Cell (RSC) of Bangladesh Road Transport Authority (BRTA) and the Police Department. Current road safety research and investigation works have been based on this database.

It is revealed that about 70 percent of road accident fatalities occur in rural areas including rural sections of national highways. Almost 80 percent of the fatalities involve vulnerable road users e.g. pedestrians, bicyclists and motorcyclists. Pedestrian-vehicle conflicts are found to be the greatest problem with significant involvement of trucks and buses. It has been observed that up to 62 percent of urban road accident deaths comprise pedestrians, and in Dhaka city, it is about 70 percent. Of the total reported accidents nearly 50 percent occur on national and regional highways. Accidents and fatalities on national highways can be characterized as clustered on some selected sections, identified as Hazardous Road Locations (HRLs). Nearly 40 percent of accidents are concentrated on around 2 percent of the highway network, demonstrating that accidents are amenable to site specific treatments. Accident type analysis shows 'hit pedestrian' as the dominant accident type both for urban and rural areas of which 45 percent resulted in fatal accidents. Other common accident types are rear end collision (16.5%), head on collision (13.2%) and overturning (9.3%). Heavy vehicles such as trucks and buses including minibuses are major contributors to road accidents (buses and minibuses 33%, trucks 27%), and in fatal accidents their shares are 35 percent and 29 percent respectively. About 2.5 percent of the reported accidents occur on bridges and culverts [Hoque *et al.*, 2010].

Various studies comprising on-site field investigations, systematic safety checks and audits, comprehensive analyses of accident reports, eyewitness and victim interviews, drivers' observations and opinion surveys, and expert opinion surveys, have been conducted by different organizations to identify the causative factors of road accidents. These studies reveal that the principal contributing factors to accidents are deficiencies in land-use and road network planning, adverse roadway and roadside environments, absence of or inappropriate pedestrian facilities, defective bridges and bridge approaches, inappropriate intersection designs, reckless driving, vehicle defects, presence of non-standard informal vehicles on main roads and unauthorized vehicle modifications. In addition, driver incompetency, road users' low level of awareness of the safety problem, and inadequate traffic law enforcement and sanctions were also among the major causes of accidents. However, it is difficult to quantify which factors are responsible for how many accidents due to the fact that a large number of contributory factors are not covered by the current accident reporting system [Mahmud *et al*., 2009].

Although MAAP based accident database constitutes the only repository for road traffic accident information of the country, its application is constrained by a number of limitations such as underreporting specially in case of lower severity, wrong transcription of Accident Report Forms (ARFs), improper recording of ARFs, etc. It is well recognized that road traffic accidents are usually under reported. Extent and spatial distortion of underreporting might cause inappropriate design of counter-measures and disproportion of resources. Even in case of recorded accidents, erroneous information can be evolved from improper transcription of ARFs. Also, improper recording of ARFs, lack of training and other demand at the accident scene induce internal inaccuracy in accident database [Alam *et al*., 2006].

In the field of transportation engineering large amounts of data may need to be handled, specially during studies on accident analysis and when general traffic accident data are heterogeneous. Moreover, in Bangladesh accident data are sometimes biased and such limitations cannot be overcome by general statistical methods. Statistics tables and ordinary charting techniques are not sufficient for present day requirements and this causes difficulties in the effective visualization of results and patterns. So, it is unrealistic to draw conclusions based on these data. Another disadvantage is that ordinary methods limit human involvement in the exploration tasks due to large sample, missing data, computational difficulty, etc.

The existing road accident analysis system in Bangladesh is more focused onto record management and basic data analysis. The road accident data are yet to be fully utilized for decision making and performance monitoring because the existing system is unable to perform extensive and detailed analysis on road safety. Accident data are often kept just for record keeping purposes rather than using it as a source of intelligence. However, most of the previous studies focused on a few risk factors, some specific road users or certain types of crashes; and so the important factors affecting injury or crash severity have not been yet completely recognized. The prerequisite to improve road safety is to have a comprehensive road accident database and analysis system. Advanced road accident analysis system is needed to help strategize road safety initiative as well as inculcate better understanding of road accident causation. Furthermore, accident data are critical to monitor and evaluate the effectiveness of road safety interventions introduced by the government and road authorities.

Advanced data analysis system has the potential to take advantage of the available accident data. Better structured data will create conditions for deeper analysis, aiding in the formulation of evidence-based research on road safety and enabling better road safety interventions as well as performance monitoring. The system will use the road accident database as the source of intelligence, to help determine accident causation and provide a clearer picture of the issues and potential intervention to improve the road safety condition. Data mining is such an approach that focuses on searching for new and interesting hypotheses than confirming the present ones. It includes various tools, techniques and applications that can be applied to eliminate the road accident data related deficiencies as well as statistical limitations. Therefore, it has been utilized for finding yet unrecognized and unsuspected facts especially in the field of road safety. This gives the basis to conduct this research.

Progress in digital data acquisition and storage technology has resulted in the growth of huge databases. This has occurred almost everywhere, from the mundane (such as supermarket transaction data, credit card usage records, telephone call details, and government statistics) to the more exotic (such as images of astronomical bodies, molecular databases, and medical records) areas of human endeavor. Little wonder, then, that interest has grown in the possibility of tapping these data, of extracting from them information that might be of value to the owner of the database. The discipline concerned with this task has become known as data mining. Defining a scientific discipline is always a controversial task; researchers often disagree about the precise range and limits of their fields of study. Bearing this in mind, and accepting that others might disagree about the details, working definition of data mining can be adopted as: ‘data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner’ [Hand *et al*., 2001].

The relationships and summaries derived through a data mining exercise are often referred to as models or patterns. Examples include linear equations, rules, clusters, graphs, tree structures, and recurrent patterns in time series. The definition above refers to observational data, as opposed to experimental data. Data mining typically deals with data that have already been collected for some purpose other than the data mining analysis. This means that the objectives of the data mining exercise play no role in the data collection strategy. This is one way in which data mining differs from much of statistics, in which data are often collected by using efficient strategies to answer specific questions. For this reason, data mining is often referred to as secondary data analysis. The definition also mentions that the data sets examined in data mining are often large. If only small data sets were involved, we would merely be discussing classical exploratory data analysis as practiced by statisticians. When we are faced with large bodies of data, new problems arise. Some of these relate to housekeeping issues of how to store or access the data, but others relate to more fundamental issues, such as how to determine the representativeness of the data, how to analyze the data in a reasonable period of time, and how to decide whether an apparent relationship is merely a chance occurrence not reflecting any underlying reality [Hand *et al*., 2001].

Ideally in statistical analysis, one designs and conducts experiments and then tests the validity of hypotheses from data collected. One gains an understanding of the properties of the data from the underlying distributions. The validity of a hypothesis is established from analyzing the distributions. In many cases, the data does not represent the outcome of a structured experiment. In such cases, methods that allow for the discovery of patterns in the data are needed. Methods for determining dominant patterns in data are usually referred to as ‘Data Mining’. Furthermore, the data from these unstructured experiments tend to be enormous. Data mining methods typically make or require assumptions in order to control computational complexity [Ekhaus, 2003].

The two approaches of learning from data or turning data into knowledge are complementary. The information obtained from a bottom-up analysis (data mining), which identifies important relations and tendencies, can not explain why these discoveries are useful and to what extent they are valid. The confirmatory tools of top-down analysis (statistics) can be used to confirm the discoveries and evaluate the quality of decisions based on those discoveries. Performing a top-down analysis, we may think of possible explanations for the observed behavior and let those hypotheses dictate the data to be analyzed. Then, performing a bottom-up analysis, we let the data suggest new hypotheses to test [Statoo Consulting, Switzerland].

In this research, an attempt has been made to study the feasibility and utility of data mining methods in the context of road traffic safety of Bangladesh. As data mining covers a large and versatile set of methods for large-scale data analysis, exploratory and descriptive methods have been emphasized in this study. The intention was to find out whether robust clustering together with association and item sets mining techniques were able to elicit reasonable, and hopefully novel, unsuspected and interesting facts from road traffic accident data.

1.2 Purpose and Objectives

The purpose of the research was to investigate the feasibility and utility of data mining methods in the context of road traffic safety in Bangladesh, using RTA data (1998-2010) from ARI, BUET. The specific objectives for this research were:

- To employ Hierarchical Clustering (HC) to form natural data groups and identify the hazardous clusters;
- To identify the high impact variables using Random Forest (RF) to facilitate calculation and reduce the complexity of the study; and
- To carry out an in-depth analysis on the hazardous clusters with Classification and Regression Tree (CART) method using the predictors determined by RF.

1.3 Scope of the Research

This study is concerned about the application of data mining in figuring multiple predictors' relationships towards accident severity. The study reveals how the accident severity is related to different predictors of accident events or which predictors trigger what kind of accident severity. However, in-depth analyses of the data mining findings required for developing countermeasures and policy level decisions were beyond the scope of this thesis.

1.4 Thesis Outline

The thesis has been organized into six chapters.

Chapter 1 has explained the background and motivation, purpose and objectives as well as the scope of the research.

Chapter 2 has been dedicated to review the relevant literature to formulate the concept of data mining in the context of this study.

Chapter 3 has illustrated the fundamentals of various methods in data mining that have been applied in this thesis. These include Hierarchical Clustering (HC), Random Forest (RF), and Classification and Regression Tree (CART). The descriptions are brief yet self-containing.

Chapter 4 has been dedicated to introduce the present road traffic accident database system of Bangladesh. It also accommodates a short preliminary statistical analysis of the data. The limitations of the present system have been highlighted as well.

Chapter 5 has addressed the detailed analysis and interpretation of results regarding data mining methodologies. The source of accident data and how it was incorporated in this study has also been discussed in this chapter.

Chapter 6 has presented the findings of the thesis along with its limitations and future scopes.

CHAPTER 2
LITERATURE REVIEW

2.1 Introduction

Data mining is quite a new addition to the ever growing efforts of transportation researchers to improve road traffic safety in Bangladesh. Being at its nascent stage, the study can still be considered as a distinct effort. This chapter commences by defining data mining from different perspectives. Later, it proceeds by clarifying the concept from transportation point of views. Then it summarizes the existing relevant literatures and thereby conducts a thorough review on the purpose, directions and progresses made in this emerging and increasingly important research field. It presents the cutting edge method data mining by systematically combining the thoughts of different researchers which helps in understanding how this thesis has contributed in both scientific and practical fields.

2.2 Data Mining

Data mining (DM) is used to discover patterns and relationships in data, with an emphasis on large observational databases. It sits at the common frontiers of several fields including database management, artificial intelligence, machine learning, pattern recognition, and data visualization. From a statistical perspective it can be viewed as computer automated exploratory data analysis of usually large complex datasets. This field is having a major impact on business, industry, and science. It also affords enormous research opportunities for new methodological developments. Despite the obvious connections between data mining and statistical data analysis, most of the methodologies used in data mining have so far originated in fields other than statistics [Friedman, 1997]. The definition of data mining largely depends on the background and views of the definer. Following are a few definitions taken from different sources [Friedman, 1997]:

From pattern recognition viewpoint: Data mining is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data – Fayyad.

From database view point: Data mining is the process of extracting previously unknown, comprehensible, and actionable information from large databases and using it to make crucial business decisions – Zekulin.

From machine learning view point: Data Mining is a set of methods used in the knowledge discovery process to distinguish previously unknown relationships and patterns within data – Ferruzza.

Data mining is the process of discovering advantageous patterns in data – John.

Data mining is a decision support process where we look in large databases for unknown and unexpected patterns of information – Parsaye.

Data mining and statistics are disciplines which are largely defined by the methods they use, rather than the problems they solve. Although their methods mostly do not overlap, both subjects deal with data. It seems pragmatic to utilize methods from any discipline that would help answer our questions. The differences are somewhat exacerbated by a certain lack of rigor among data mining methodologies, at least from the view point of statistics. Alternatively, traditional statistical methods do not handle the data volumes that data mining typically deals with. Today’s computers store enormous volumes of data and the rate at which it is growing is ever increasing. It is reasonable to expect new computational methods to be developed to address these growing needs. Data mining is useful for discovering relationships and statistics is useful for analyzing relationships. The two disciplines need to coexist and methods that bridges the gaps between the two are needed [Ekhaus, 2003].

2.3 Data Mining in Transportation Engineering

In the field of transportation engineering large volume of data are generated during the studies of traffic management, accident analysis, pavement conditions, roadway feature inventory, traffic signals and signal inventory, bridge maintenance, road characteristics inventory, etc. Based on these data decision makers arrive at a decision to solve a respective problem. Decision makers are always on lookout for ways to ease the pain in obtaining access to and applying disparate datasets. The basic requirements include the ability to identify what data are available, determine the characteristics of the data, extract the data of interest, and transform the data into formats necessary for applications. In real life situation of transportation domain, diverse fields of data need to be collected to integrate and to arrive at solutions. Data mining approaches have opened a new horizon for decision makers in transportation engineering [Barai, 2003].

There is a broad spectrum of engineering problems where computational intelligence is becoming an essential part in many advanced systems. Hence new techniques for extracting important knowledge from raw data are required to handle the components efficiently. Data mining is a step in this knowledge process. Basic steps of data mining and knowledge discovery are depicted in Figure 2.1 [Barai, 2003]. Detailed explanation can be found in Fayyad *et al.*, 1996.

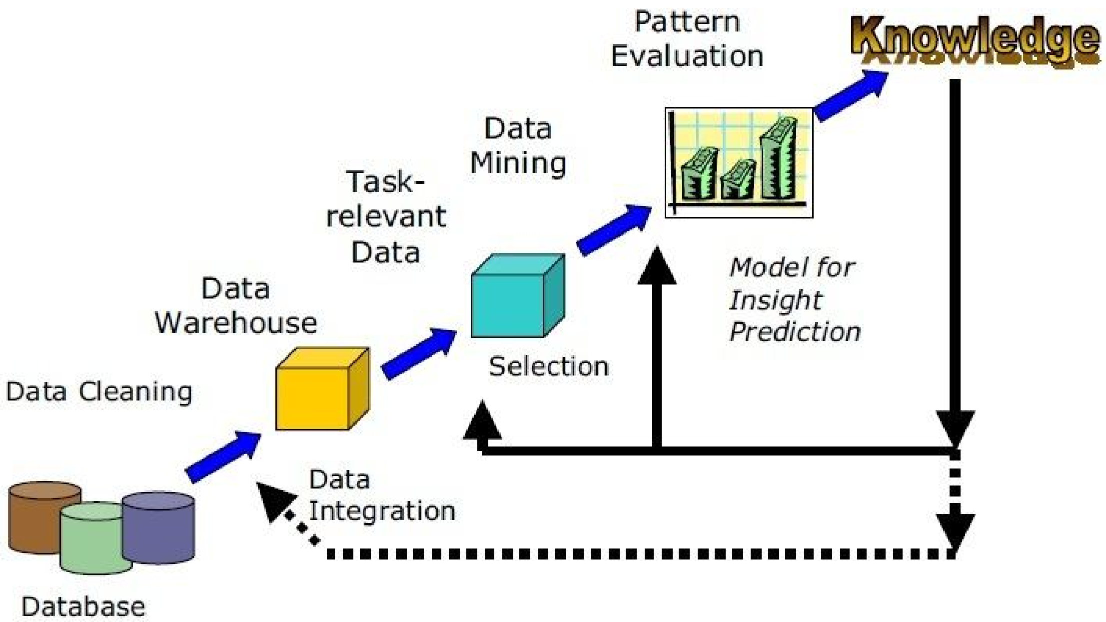


Figure 2.1. Data mining and knowledge discovery process.

2.4 Relevant Studies

In the late 90's and the beginning of this century there have been several attempts to use data mining techniques in the area of traffic safety. In particular, frequent patterns in accident data have been searched by implementing spatial data mining [Zeitouni and Chelghoum, 2001], clustering techniques [Ljubic *et al*, 2002; Geurts *et al*., 2003; Bayam *et al*., 2005], rule induction [Geurts *et al*., 2003; Geurts *et al*., 2005, Kavsek *et al*., 2006], decision trees [Strnad *et al*., 1998; Clarke *et al*., 1998; Bayam *et al*., 2005] and neural networks [Mussone *et al*., 1999; Bayam *et al*., 2005]. Some applications have combined data mining techniques with technological enhancements [Ng *et al*., 2002], for example a combination of cluster analysis, regression analysis and geographical information system (GIS) platforms to group homogeneous accident data, to estimate the number of accidents and to assess the crash risk.

In recent years there has been a growing body of research exploring whether data mining techniques are potentially more suitable than classical econometric models to uncover relations between the variables that affect accidents, such as road characteristics, driver characteristics and attitudes, vehicle features and seasonal factors. Clustering methods seemed an important tool when analyzing traffic accidents as these methods are able to identify groups of road users, vehicles and road segments which would be suitable targets for countermeasures [Cameron, 1997]. Lee *et al*. (2002) presented a review and discussed limitations of classical econometric models that had been widely used to analyze road crashes. Chen and Jovanis (2002) showed that certain problems might arise when using classic statistical analysis on datasets with large dimensions, namely the exponential increase in the number of parameters as the number of variables increases and the invalidity of statistical tests as a consequence of sparse data in large contingency tables. Chang and Chen (2005) compared prediction performances of decision trees and negative binomial regressions to determine that decision trees were a better method for analyzing freeway accident frequencies. Chong *et al*. (2005) evaluated the performance of four machine learning paradigms applied to modeling the severity of injury that occurred during traffic accidents: neural networks, support vector machines, decision trees and a hybrid model involving decision trees and neural networks.

Kim (1995) developed a log-linear model to clarify the role of driver characteristics and behaviors in the causal sequence leading to more severe injuries. They found that driver behaviors of alcohol or drug use and lack of seat belt use greatly increase the odds of more severe crashes and injuries.

Shankar (1996) applied a nested logic formulation for estimating accident severity likelihood conditioned on the occurrence of an accident. The study found that there is a greater probability of evident injury or disabling injury/fatality relative to no evident injury if at least one driver did not use a restraint system at the time of the accident.

Dia (1997) used real-world data for developing a multilayered NN freeway incident detection model. They compared the performance of the neural network model and the incident detection model in operation on freeways.

Abdalla *et al*. (1997) also studied the relationship between casualty frequency and the distance of an accident from residential zones. Not surprisingly, casualty frequencies were higher in accidents that occurred nearer to residential zones, possibly due to higher exposure. The casualty rates among residents from relatively deprived areas were significantly higher than those from relatively affluent areas.

Yang (1999) used NN approach to detect safer driving patterns that have less chances of causing death and injury when a car crash occurs. Evanco (1999) conducted a multivariate population-based statistical analysis to determine the relationship between fatalities and accident notification times. The analysis demonstrated that accident notification time is an important determinant of the number of fatalities for accidents on rural roadways.

Mussone *et al*. (1999) used neural networks to analyze vehicle accidents that occurred at intersections in Milan, Italy. They used feed-forward multilayer perception (MLP) with BP learning. The model had 10 input nodes for eight variables: day/night, traffic flows in the intersection, number of virtual and real conflict points, intersection type, accident type, road surface condition, and weather condition. The output node (accident index) was calculated as the ratio between the number of accidents at a given intersection and at the most dangerous intersection. Results showed that the highest accident index for the running over of pedestrians occurred at non-signalized intersections at nighttime.

Ossenbruggen *et al*. (2001) used a logistic regression model to identify the prediction factors of crashes and crash-related injuries, using models to perform a risk assessment of a given region. These models included attributes describing a site by its land use activity, roadside design, use of traffic control devices, and traffic exposure. Their study illustrated that village sites were less hazardous than residential or shopping sites.

Sohn and Hyungwon (2001) conducted research on pattern recognition in the framework of RTA severity in Korea. They observed that an accurately estimated classification model for several RTA severity types as a function of related factors provided crucial information for accident prevention. Their research used three data mining techniques, neural network, logistic regression, and decision tree, to select a set of influential factors and to construct classification models for accident severity. Their three approaches were then compared in terms of classification accuracy. They found that accuracy did not differ significantly for each model, and that the protective device was the most important factor in the accident severity variation.

Bedard (2002) applied a multivariate logistic regression to determine the independent contribution of driver, crash, and vehicle characteristics to drivers' fatality risk. It was found that increasing seatbelt use, reducing speed, and reducing the number and severity of driver side impacts might prevent fatalities.

Ossiander (2002) used Poisson regression to analyze the association between the fatal crash rate (fatal crashes per vehicle mile traveled) and the speed limit increase and found that the speed limit increase was associated with a higher fatal crash rate and more deaths on freeways.

To analyze the relationship between RTA severity and driving environment factors, Sohn and Lee (2002) used various algorithms to improve the accuracy of individual classifiers for two RTA severity categories. Using neural network and decision tree individual classifiers, three different approaches were applied: classifier fusion based on the Dempster–Shafer algorithm, the Bayesian procedure, and logistic model; data ensemble fusion based on arcing and bagging; and clustering based on the k-means algorithm. Their empirical results indicated that a clustering-based classification algorithm works best for road traffic accident classification in Korea.

Ng, Hung and Wong (2002) used a combination of cluster analysis, regression analysis, and geographical information system (GIS) techniques to group homogeneous accident data, estimate the number of traffic accidents, and assess RTA risk in Hong Kong. Their resulting algorithm displayed improved accident risk estimation compared to estimates based on historical accident records alone. The algorithm was more efficient, especially for fatality and pedestrian related accident analyses. The authors claimed that the proposed algorithm could be used to help authorities effectively identify areas with high accident risk, and serve as a reference for town planners considering road safety.

Chang and Chen (2005) conducted data mining research focusing on building tree-based models to analyze freeway accident frequency. Using the 2001- 2002 accident data of National Freeway 1 in Taiwan, the authors developed classification and regression tree (CART) and negative binomial regression models to establish the empirical relationship between traffic accidents and highway geometric variables, traffic characteristics, and environmental factors. CART is a powerful tool that does not require any pre-defined underlying relationship between targets (dependent variables) and predictors (independent variables). The authors found that the average daily traffic volume and precipitation variables were the key determinants of freeway accident frequency. Furthermore, a comparison of their two models demonstrated that CART is a good alternative method for analyzing freeway accident frequencies.

Beshah (2005) analyzed historical RTA data, including 4,658 accident records at the Addis Ababa Traffic Office, to investigate the application of data mining technology to the analysis of accident severity in Addis Ababa, Ethiopia. Using the decision tree technique and applying the Knowledge SEEKER algorithm of the Knowledge STUDIO data mining tool, the developed model classified accident severity into four classes: fatal injury, serious injury, slight injury, and property damage. Accident cause, accident type, road condition, vehicle type, light condition, road surface type, and driver age were the basic determinant variables for injury severity level. The classification accuracy of this decision tree classifier was reported to be 87.47 percent.

Chang and Wang (2006) applied non-parametric classification tree techniques to analyze accident data from the year 2001 for Taipei, Taiwan. A CART model was developed to establish the relationship between injury severity and driver/vehicle characteristics, highway/environment variables, and accident variables. The most important variable associated with crash severity was the vehicle type, with pedestrians, motor-cyclers, and bicyclists having the highest injury risks of all driver types in the RTAs.

Using one clustering (Simple K-Means) and three classification (J48, naïve Bayes, and One R) algorithms, Srisuriyachai (2007) analyzed road traffic accidents in the Nakhon Pathom province of Bangkok. Considering the descriptive nature of the results and classification performance, the J48 algorithm was sufficiently useful and reliable. The outcome of the research was traffic accident profiles, which the author presented as a useful tool for evaluating RTAs in Nakhon Pathom.

Wong and Chung (2008) used a comparison of methodology approaches to identify causal factors of accident severity. Accident data were first analyzed with rough set theories to determine whether they included complete information about the circumstances of their occurrence according to an accident database. Derived circumstances were then compared. For those remaining accidents without sufficient information, logistic regression models were employed to investigate possible associations. Adopting the 2005 Taiwan single-auto-vehicle accident data set, the results indicated that accident fatality resulted from a combination of unfavorable factors, rather than from a single factor. Moreover, accidents related to rules with high or low support showed distinct features.

Following Beshah's (2005) work, Zelalem (2009) conducted a data mining study to classify driver responsibility levels in traffic accidents in Addis Ababa. The study focused on identifying the important factors influencing the level of driver responsibility, and used the RTA dataset of the Addis Ababa Traffic Control and Investigation Department (AATCID). The WEKA data mining tool was used to build the decision tree (using the ID3 and J48 algorithms) and MLP (back propagation algorithm) predictive models. Rules representing patterns in the accident dataset were extracted from the decision tree, revealing important relationships between variables influencing a driver's level of responsibility (e.g., age, license grade, education, driving experience, and other environmental factors). The accuracies of these models were

88.24% and 91.84%, respectively, with the decision tree model found to be more appropriate for the problem type under consideration.

Getnet (2009) investigated the potential application of data mining tools to develop models supporting the identification and prediction of major driver and vehicle risk factors that cause RTAs. The research used the WEKA version 3-5-8 tool to build the decision tree (using the J48 algorithm) and rule induction (using PART algorithm) techniques. Performance of the J48 algorithm was slightly better than that of the PART algorithm. The license grade, vehicle service year, vehicle type, and experience were identified as the most important variables for predicting accident severity.

Liu (2009) developed a decision support tool for liability authentications of two-vehicle crashes, based on self-organizing feature maps (SOM) and data mining models. Although the study used a small data sample, the decision support system provided reasonably good liability attributions and references on the given cases.

2.5 Recent Advancements

Researchers over the past two decades have conducted significant number of studies to identify factors influencing crash [Fridstrom *et al*., 1995; Miaou and Song, 2005] and developed crash prediction models to calculate the frequency and associated severity of crash on conventional expressways [Khan *et al*., 1999; Caliendo *et al*., 2007]. Several analogous studies have underscored positive correlations between traffic flow variables and road crashes [Cedar and Livneh, 1982; Cedar, 1982; Frantzeskakis and Iordanis, 1987] that brought long-term safety benefits by improving geometric designs, road side environment and helping in decision making for budget allocation, albeit the countermeasures were rather reactive in nature [Oh *et al*., 2001; Lee *et al*., 2003]. They also ignored the complex interaction among traffic flow variables that may have abetted crashes. This is as they employed highly aggregated traffic data (e.g., hourly, daily or yearly flow) which could not capture the suddenly developed hazardous traffic conditions that could lead to a road crash [Hossain and Muromachi, 2013].

Recently, with the enhanced data collection, storage and analysis capabilities, researchers have started paying attention in developing proactive road safety management systems for expressways using high-resolution real time traffic data. Several real-time crash prediction models have been proposed based on the hypothesis that the probability of a crash on a specific road section can be predicted for a very short time window using the instantaneous traffic flow data [Lee *et al*., 2002, 2003; Golob *et al*., 2003; Pande and Abdel-Aty, 2005]. This opened the possibility to develop proactive road safety management systems which may even be able to prevent some crashes that would have taken place otherwise [Lee *et al*., 2002, 2003; Abdel-Aty and Pande, 2004; Abdel-Aty and Abdalla, 2004; Oh *et al*., 2005a,b; Abdel-Aty *et al*., 2006a,b; Dias *et al*., 2009; Hossain and Muromachi, 2010b].

Jang *et al*. (2012) extended the study horizon by introducing a real-time collision warning system for the intersections where conditions related to vulnerable line of site and/or traffic violation can be observed. Christoforou *et al.* (2012) in their studies have determined crash probability along with associated crash severity. However, these studies were focused on improving the prediction capability rather than providing insight into crash phenomena. Among the studies related to identifying the traffic variables leading to crash, Abdel-Aty *et al*. (2005) ascertained that crashes occur in high speed and low speed scenarios. While the former is caused by quick formation and subsequent dissipation of queues causing a backward shock wave, the latter is due to a disruption in the downstream that propagates a shock wave to the upstream impending driving errors.

With a similar approach but including only rear-end crash data, Pande and Abdel-Aty (2006a) affirmed that crashes are related to coefficient of variation in speed and average occupancy under extended congestion. They also found that the high speed crashes were more explainable with average speed and occupancy in a downstream detector. They mentioned that presence of ramp in the downstream have impact on crash but did not shed light on the types of ramps and their relative vicinity. Two simultaneous studies were conducted on the same study area (I-4, Ontario, FL, USA) for lane-changing related collisions and it was found that average speeds at upstream and downstream together with difference in occupancy on adjacent lanes and standard deviation of volume and speed at a location downstream of the crash point are the major contributing factors [Pande and Abdel-Aty, 2006b; Lee *et al*., 2006].

Dias *et al*. (2009) introduced level of congestion rather than the aggregated speed of vehicles as a predictor and affirmed a positive correlation between congestion and crash risk. Zheng *et al*. (2010) considered only congested traffic condition and used matched control logistic regression to prove that traffic oscillations contribute to crash. Christoforou *et al*. (2011) utilized real-time traffic data to associate different traffic parameters with various crash types. Xu *et al*. (2012) suggested that traffic characteristics leading to crash vary substantially between congested and uncongested situations.

The studies existing were more concerned about identifying the factors and placed little or no concentration on why and how these factors contribute to a crash. They in most cases did not verify if the factors vary for the basic freeway segments (BFS) and ramp areas. McCartt *et al*. (2002) found different crash types and characteristics dominating different types of ramps. Chen *et al*. (2009, 2010) found significant safety impact even for off ramps of freeways when they had different number and arrangements of lanes. Due to high variation in ramp density between conventional expressways and urban expressways, the relevance and transferability of the findings of these studies to urban expressways may not be justified adequately. Thus, it was important to investigate if the existing findings were generic to all kinds of expressways or whether they differ significantly [Hossain and Muromachi, 2013].

Hossain and Muromachi (2013) in their study employed high resolution detector data to identify the traffic patterns impending hazardous driving conditions. Unlike the previous studies, their study separated the road sections of the urban expressways into five groups – the basic freeway segments (BFS) and areas near downstream (d/s) and upstream (u/s) of the on (entrance) and off (exit) ramps and attempts to identify generic crash prone traffic patterns for each of these groups. They came up with the fact that the high risk clusters in all the five groups of the road sections had substantially high differences in their congestion indexes which indicated either the downstream or the upstream traffic conditions were at least partially congested. Thus, it was easier to explain the crash mechanism under low speed operation. This was also logical to believe that many high speed crashes might be associated with unsafe driving rather than traffic condition which was hazardous and thus hard to explain with traffic flow variables. Therefore, education and enforcement related interventions are required as well.

2.6 Summary

Relevant literatures have highlighted enormous scopes regarding the application of data mining (DM) on road traffic accident database. The studies have outlined that DM has the potential to quantify multiple predictors' relationships towards accident instances. DM is such an approach that focuses on searching for new and interesting hypotheses than confirming the present ones. It includes various tools, techniques and applications that can be applied to eliminate the road accident data related deficiencies as well as statistical limitations. Therefore, it has been utilized for finding yet unrecognized and unsuspected facts especially in the field of road safety.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Introduction

This study involves handling large variable spaces of accident predictors, clustering them into major sub-groups, finding their relative importance, understanding their interaction that can predict the underlying factors of accident severities. To achieve these, representing knowledge properly and making decisions based on the data are very important. Hence, several data mining methods have been employed in this research along with general statistical tables and graphs. This chapter provides a brief but self-containing description of these methods along with their applicability. The chapter also elaborates the data collection and data preparation processes for this research.

3.2 Methods and Work Flow of the Study

The analytical part of this thesis can be separated broadly into two phases – understanding present road safety status of the country, and applying data mining to come up with some novel, unsuspected, and reasonable facts from road traffic accident data. The first phase comprises with general statistical analysis i.e. generating tables and graphs through SQL at MS Access using ARI's accident database. It also outlined the present analytical practice of RTA data in Bangladesh. The second phase is the respiratory part of this research. Three data mining methods have been applied for this phase. At first, hierarchical clustering methodology was employed to form natural data groups and to identify hazardous clusters; then random forest was applied to identify, rank, and thus select a subset of variables from a large variable space, to be considered for this study. Finally, classification and regression trees have been allowed to investigate the accident severity mechanism of the hazardous clusters. Following sections describe the methods sequentially.

3.3 Hierarchical Clustering (HC)

Selection of a clustering methodology depends on data type. The data used in cluster analysis can be categorical/nominal (e.g. name/category i.e. data cannot be added, subtracted, multiplied or divided), ratio (data can be added, subtracted, multiplied or divided), interval (difference meaningful but cannot be multiplied or divided), and ordinal (e.g. good, very good, excellent). However, having a mixture of different types of variable makes the analysis more complicated. This is because in cluster analysis we need to have some way of measuring the distance between observations, and the type of measure used will depend on what type of data we have. Accident data is usually mixed type i.e. a single accident event is recorded with different types (categorical, ratio, interval and ordinal) of variable, and essentially in this research, the mixed attribute type is being considered [http://cran.r-project.org/web/packages/cluster/cluster.pdf].

There are a number of different methods that can be used to carry out a cluster analysis. The main reason for having many clustering methods is the fact that the notion of 'cluster' is not precisely defined [Estivill-Castro, 2000]. Consequently many clustering methods have been developed, each of which uses a different induction principle. Farley and Raftery (1998) suggest dividing the clustering methods into two main groups: hierarchical and partitioning (non-hierarchical e.g. k-means, expectation maximization) methods. Han and Kamber (2001) suggest categorizing the methods into additional three main categories: density-based methods, model-based clustering and grid-based methods. An alternative categorization based on the induction principle of the various clustering methods is presented in [Estivill-Castro, 2000]. Each clustering method has its own advantages and disadvantages. However, for mixed attribute type, HC is preferred in researcher community.

HC constructs the clusters by recursively partitioning the instances in either a top-down or bottom-up fashion. These methods can be subdivided as follows [Internet Links]:

> Agglomerative hierarchical clustering: each object initially represents a cluster of its own i.e. subjects start in their own separate cluster. The two 'closest' (most similar) clusters are then combined and this is done repeatedly until all subjects are in one cluster. Finally, the desired cluster structure is derived.

Divisive hierarchical clustering: all objects initially belong to one cluster. Then the cluster is divided into sub-clusters, which are successively divided into their own sub-clusters (i.e. the previous strategy is applied but in reverse order). This process continues until the desired cluster structure is obtained.

However, agglomerative methods are used more often than divisive methods, so this dissertation will concentrate on the former rather than the latter. The result of the hierarchical methods is a dendrogram, representing the nested grouping of objects and similarity levels at which groupings change. A clustering of the data objects is obtained by cutting the dendrogram at the desired similarity level.

Merging or division of clusters is performed according to some similarity measure, chosen so as to optimize some criterion (such as a sum of squares). HC methods could be further divided according to the manner that the similarity measure is calculated [Jain *et al*., 1999]. These methods are elucidated in the following [Internet Links]:

Single-link clustering (also called the connectedness, the minimum method or the nearest neighbor method): in this method the distance between two clusters is defined to be the distance between the two closest members, or neighbors (Figure 3.1). This method is relatively simple but is often criticized because it does not take account of cluster structure and can result in a problem called chaining whereby clusters end up being long and straggly. However, it is better than the other methods when the natural clusters are not spherical or elliptical in shape. Interested readers are requested to consult Sneath and Sokal (1973).

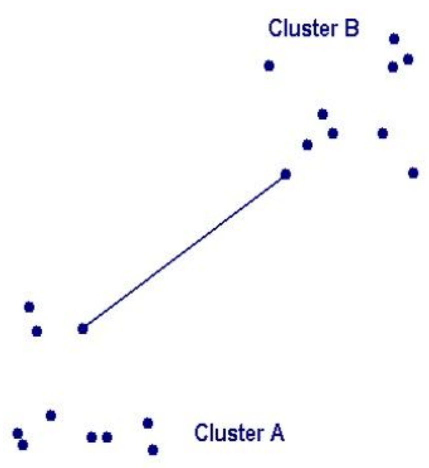


Figure 3.1. Single-link clustering.

Complete-link clustering (also called the diameter, the maximum method or the furthest neighbor method): in this case the distance between two clusters is defined to be the maximum distance between members — i.e. the distance between the two subjects that are furthest apart (Figure 3.2). This method tends to produce compact clusters of similar size but, as for the nearest neighbor method, does not take account of cluster structure. It is also quite sensitive to outliers. Interested readers are requested to consult King (1967).

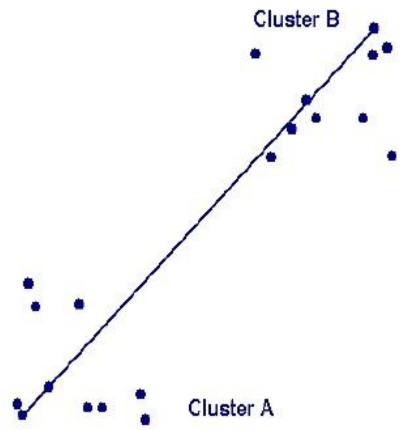


Figure 3.2. Complete-link clustering.

Average-link clustering (also called minimum variance method, sometimes referred to as UPGMA): in this method the distance between two clusters is calculated as the average distance between all pairs of subjects in the two clusters (Figure 3.3). This is considered to be a fairly robust method. Interested readers are requested to consult Ward (1963) and Murtagh (1984).

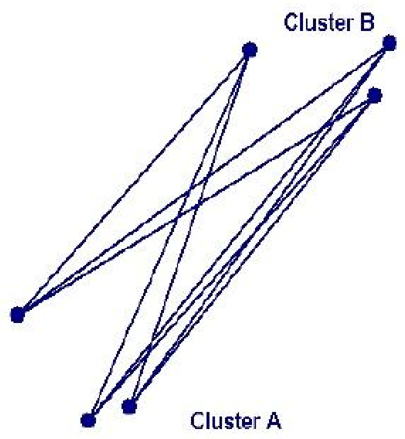


Figure 3.3. Average-link clustering.

Centroid method: here the centroid (mean value for each variable) of each cluster is calculated and the distance between centroids is used. Clusters whose centroids are closest together are merged. This method is also fairly robust.

Ward's method: in this method all possible pairs of clusters are combined and the sum of the squared distances within each cluster is calculated. This is then summed over all clusters. The combination that gives the lowest sum of squares is chosen. This method tends to produce clusters of approximately equal size, which is not always desirable. It is also quite sensitive to outliers. Despite this, it is one of the most popular methods.

All the above mentioned methods have their own advantages and disadvantages. Interested readers are requested to consult Guha *et al*. (1998). Considering all the options, the complete linkage method has been adopted for this thesis.

The complete-link hierarchical clustering method is exemplified for clear understanding in the following [Source: http://www.econ.upf.edu/~michael/stanford/maeb7.pdf]:

Let us consider Table 3.1 as the desired dissimilarity (distance) matrix.

Table 3.1. Dissimilarity matrix

samples	A	B	C	D	E	F	G
A	0	0.5000	0.4286	1.0000	0.2500	0.6250	0.3750
B	0.5000	0	0.7143	0.8333	0.6667	0.2000	0.7778
C	0.4286	0.7143	0	1.0000	0.4286	0.6667	0.3333
D	1.0000	0.8333	1.0000	0	1.0000	0.8000	0.8571
E	0.2500	0.6667	0.4286	1.0000	0	0.7778	0.3750
F	0.6250	0.2000	0.6667	0.8000	0.7778	0	0.7500
G	0.3750	0.7778	0.3333	0.8571	0.3750	0.7500	0

The first step in the hierarchical clustering process is to look for the pair of samples that are the most similar and closest in the sense of having the lowest dissimilarity – this is the pair B and F (Table 3.1), with dissimilarity equal to 0.2000. These two samples are then joined at a level of 0.2000 in the first step of the dendrogram, or clustering tree (see first part of Figure 3.4, and the vertical scale of 0 to 1 which calibrates the level of clustering). The point at which they are joined is called a node.

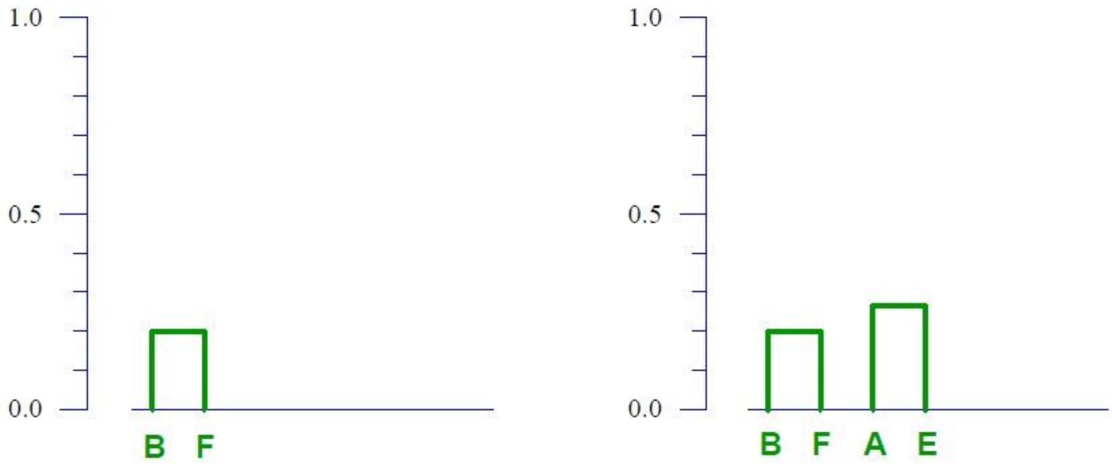


Figure 3.4. First step in dendrogram.

This step has been repeated, but the problem remains how to calculate the dissimilarity between the merged pair (B,F), and the other samples. This decision is dependent on the type of hierarchical clustering intended to perform, and there are several choices. For the moment, one of the most popular ones is chosen, called the maximum or complete linkage method - the dissimilarity between the merged pair and the others will be the maximum of the pair of dissimilarities in each case. For example, the dissimilarity between B and A is 0.5000, while the dissimilarity between F and A is 0.6250. Hence the maximum of the two, 0.6250, is chosen to quantify the dissimilarity between the merged pair (B,F) and A. Thus a new dissimilarity matrix is attained (Table 3.2).

Table 3.2. Dissimilarity matrix after first merging

samples	A	(B,F)	C	D	E	G
A	0	0.6250	0.4286	1.0000	0.2500	0.3750
(B,F)	0.6250	0	0.7143	0.8333	0.7778	0.7778
C	0.4286	0.7143	0	1.0000	0.4286	0.3333
D	1.0000	0.8333	1.0000	0	1.0000	0.8571
E	0.2500	0.7778	0.4286	1.0000	0	0.3750
G	0.3750	0.7778	0.3333	0.8571	0.3750	0

The process is now repeated: finding the smallest dissimilarity in Table 3.2, which is 0.2500 for samples A and E, and then cluster these at a level of 0.25, as shown in the second part of Figure 3.4. Then recomputed the dissimilarities between the merged pair (A,E) and the rest to obtain Table 3.3. For example, the dissimilarity between the merged pairs (A,E) and (B,F), is the maximum of 0.6250 (A to (B,F)) and 0.7778 (E to (B,F)).

Table 3.3. Dissimilarity matrix after second merging

samples	(A,E)	(B,F)	C	D	G
(A,E)	0	0.7778	0.4286	1.0000	0.3750
(B,F)	0.7778	0	0.7143	0.8333	0.7778
C	0.4286	0.7143	0	1.0000	0.3333
D	1.0000	0.8333	1.0000	0	0.8571
G	0.3750	0.7778	0.3333	0.8571	0

In the next step the lowest dissimilarity in Table 3.3 is 0.3333, for C and G – these are merged, as shown in the first diagram of Figure 3.5, to obtain Table 3.4. Now the smallest dissimilarity is 0.4286, between the pairs, (A,E) and (B,G), and they are shown merged in the second diagram of Figure 3.5. Table 3.5 shows the last two dissimilarity matrices in this process, and Figure 3.6 the final two steps of the construction of the dendrogram, also called a binary tree because at each step two objects (or clusters of objects) are merged. As 7 objects are to be clustered in this case, there are 6 steps in the sequential process (i.e. one less) to arrive at the final tree where all objects are in a single cluster. The botanists may consider this is as an upside-down tree.

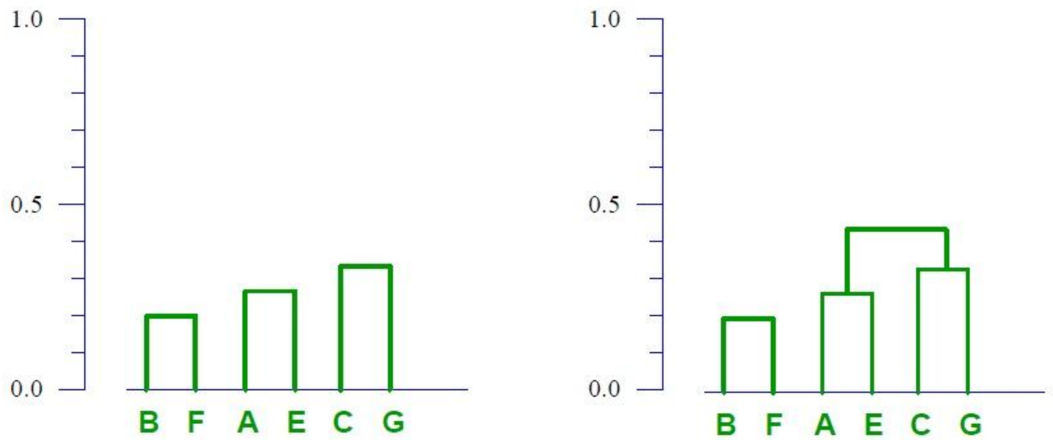


Figure 3.5. Second step in dendrogram.

Table 3.4. Dissimilarity matrix after third merging

samples	(A,E)	(B,F)	(C,G)	D
(A,E)	0	0.7778	0.4286	1.0000
(B,F)	0.7778	0	0.7778	0.8333
(C,G)	0.4286	0.7778	0	1.0000
D	1.0000	0.8333	1.0000	0

Table 3.5. Dissimilarity matrices in last step

samples	(A,E,C,G)	(B,F)	D
(A,E,C,G)	0	0.7778	1.0000
(B,F)	0.7778	0	0.8333
D	1.0000	0.8333	0

samples	(A,E,C,G,B,F)	D
(A,E,C,G,B,F)	0	1.0000
D	1.0000	0

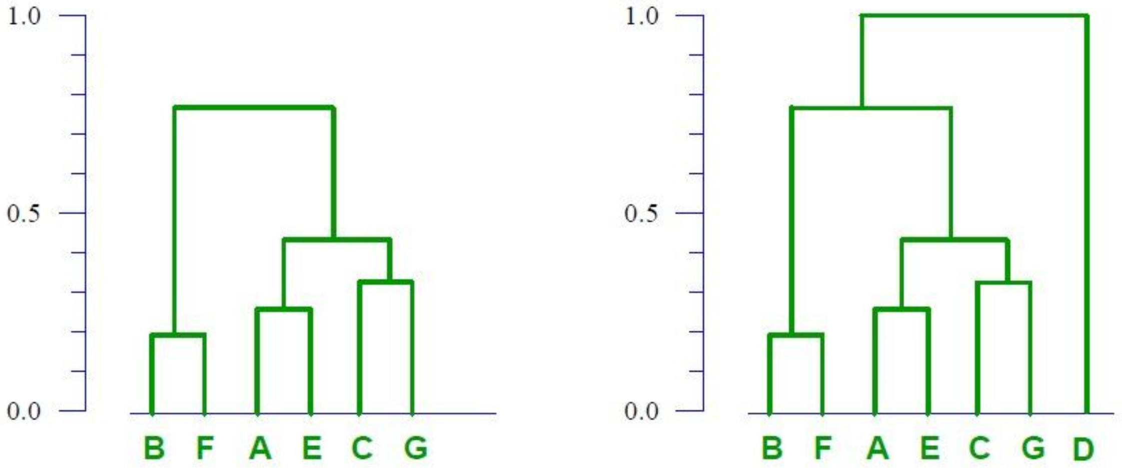


Figure 3.6. Final step in dendrogram.

The dendrogram on the right side of Figure 3.6 is the final result of the cluster analysis. In the clustering of n objects, there are n–1 nodes (i.e. 6 nodes in this case).

This study uses cluster package, daisy function, gower metric, hclust function, and cutree function of the R program to form the dissimilarity matrix, to perform the hierarchical clustering, to construct the dendrogram and to cut the tree to an appropriate size.

3.4 Random Forest (RF)

Random forest (RF) is one of the new methods in ensemble learning that can perform classification and regression as well as numerically rank the importance of the predictors in the model. Currently, RF is considered as one of the latest and most efficient methods in evaluating and ranking variable importance [Harb *et al.*, 2009]. It has demonstrated high capability in handling multicolinearity issue of large feature spaces by using two well-known methods in ensemble learning that are applied in classification trees – boosting [Shapire *et al.*, 1998] and bagging [Breiman, 1996] coupled with the idea of random variable selection. In case of boosting, the successive trees associate extra weight to points misclassified by earlier predictors. Finally, a weighted vote is taken for prediction. Whereas in bagging the earlier trees do not influence the successive trees and each is independently constructed based on a bootstrap sample (bootstrapping constructs a number of re-samples of the original dataset, each equal to the size of the original dataset, where each re-sample is produced by random sampling with replacement from the original dataset) of the dataset. Lastly, prediction is performed by conducting a simple majority voting [Liaw and Wiener, 2002]. RF adds an additional layer of randomness to bagging. To elaborate more, RF generates a given number of CART trees with a different bootstrap sampling for each tree. However, it differs slightly in the process of growing the tree through splitting. Instead of finding the best splitter at each node from all the available variables, it calculates the best splitter from a subset of variables randomly chosen from complete variables space [Hossain, 2011]. The study employed 'random forest' package of R program to implement random forest.

The major steps of the RF algorithm are [Hossain, 2011]:

(i) Let L be the complete dataset with M predictors and N records and B the total number of CART trees in the RF. Let L_b be the b-th bootstrap sample created by randomly selected n samples with replacement from L. Rest of the data, i.e., L-L_b, are called the out of bag data (OOB) of b-th bootstrap sample.

(ii) Next, for the b-th tree T_b, instead of growing a CART tree with M predictors, m predictors are randomly selected from M predictor space ($M > m$) at every node and the best splitter among *m* capable of producing two maximum pure nodes is used to split the node at each level.

(iii) Predicting from new data: run down the new data through each and every (here *B* number of trees) tree and the class of the new data is the class of the leaf of each tree where it ended up. The final class of the data is calculated by aggregating the predictions of the *B* trees. In case of classification trees, it is achieved by majority voting.

(iv) Estimating OOB error rate: at each and every bootstrap iteration the L-L_b datasets are used to calculate the misclassification rate r_b of tree T_b (this misclassification rate r_b is used for calculating the variable importance as well). This is achieved by running down the L-L_b dataset into T_b grown in step (ii). The class of each of the data points are decided based on majority voting (can be weighted). This majority voting is required only for estimating the OOB error rate (not for variable importance). In another way it can be said that lastly the r_b of all the *B* trees are aggregated to calculate the OOB error rate.

(v) Variable importance: the idea of variable importance in RF differs from conventional statistical approaches. Here, it is measured by permuting the values of each variable (one variable at a time) and then calculating the new error rate. The permuted variable with the highest error rate is considered as the most important variable as any error in measuring its value has the highest impact on the classification performance of RF. Thus, the values of the *j*-th predictor of *M* predictors in L-L_b are permuted and the new dataset is used to calculate the misclassification rate r^j_b. Here, $|r_b - r^j_b|$ is the variable importance V_j of the *j*-th variable in the *b*-th tree. The process is repeated for *B* trees and the final variable importance is calculated by averaging the V_j of each variable (j = 1 to *M*).

The study employed 'randomForest' package of R program [Dalgaard, 2008] to implement random forest. Interested readers are requested to consult Breiman (2001) to acquire in-depth knowledge on random forest.

3.5 Classification and Regression Tree (CART)

Classification and Regression Tree (CART) is a method of generating decision trees developed by Breiman *et al*. (1984) that can be applied for knowledge discovery and classifying new data. In case of problem domains with large feature space, it may not be wise to opt for a global single predictive linear or polynomial regression model for the entire data space. On the contrary, CART is nonparametric by nature and partitions the data space into subdivisions in a recursive manner and brings it down to small manageable chunks containing data of only one dominant class. Its tree type structure is specially helpful to gain insight about the problem domain and facilitates identifying the most important predictors, too. The methodology has three major activities. First, it grows a decision tree of maximum depth in such a way that each end node, often referred as leaf, contains data of a pure class. The second step prunes the tree to an appropriate size and obtains a sequence of nested sub-trees. Lastly, the best classification tree is chosen and the model is ready for classifying new data [Hossain, 2011]. Although there are many algorithms available for the job, this research will explain Gini splitting rule to split the nodes and cross validation to prune the trees as the software will be used in this study uses these methods (rpart package of R program).

Let the learning dataset have M number of predictors x_i, where i = 1 to M. Let t_p be a parent node and t_l, t_r the left and the right child nodes after splitting. In CART, the splitting rule aims to separate the data into two chunks with maximum homogeneity. The algorithm ascertains the splitting value x_i^R in such a way that for all splitting values of all the variables, x_i^R ensures maximum homogeneity of the child nodes. This is calculated by defining an impurity function $I(t)$. The idea accents that x_i^R will maximize the difference between the impurity of the parent node and the child nodes as presented in Equation 3.1 [Hossain, 2011]:

$$\arg\max\,[\Delta I(t) = I(t_p) - P_l * I(t_l) - P_r * I(t_r)] \qquad (3.1)$$

where P_l and P_r are the proportions of data in left and right nodes. Several algorithms are available for defining the impurity functions that can satisfy Equation 3.1 to find the appropriate value of x_i^R.

However, it has been ascertained that the final tree is insensitive to the algorithm selected. This study adopts Gini index based splitting algorithm for node splitting. If the outcome variable has K number of categories then the Gini index will vary between zero and (1-1/K). The minimum value is observed when a node is pure, i.e., data of one class only and the maximum value is yielded when the outcome classes are equally distributed in the node. Gini index at any node t can be defined as [Hossain, 2011]:

$$I(t)=\sum_{j\neq l} p(j|t)\,p(l|t)=\sum_{j} p(j|t)(1-p(j|t))=\sum_{j} p(j|t)-\sum_{j} p(j|t)^2=1-\sum_{j} p(j|t)^2 \quad (3.2)$$

where j and l are the categories of the outcome variable and $p(j|t)$ is the proportion of outcome class j in node t. Now, the change in impurity can be calculated by plugging Equation 3.2 into Equation 3.1. The change in impurity can be maximized by minimizing $[P_l{*}I(t_l) + P_r{*}I(t_r)]$. Using this splitting algorithm, tree is grown up to the maximum depth through recursive splitting until every node contains a pure class. Subsequently, the tree is pruned through a trade off between the complexity of the tree and the misclassification error. It is achieved by minimizing a compound function called cost-complexity (cp) function as shown in Equation 3.3.

$$\min R_{\alpha}(T) = R(T) + \alpha(T') \quad (3.3)$$

where $R(T)$ is the misclassification error of tree T; T' is the total sum of terminal nodes in the tree T and $\alpha(T')$ is the complexity measure. The cross-validation method calculates the value of α by repeatedly taking a part of the data as learning sample to build the tree and using the other part to test the classification accuracy [Hossain, 2011].

The value of α can be calculated in many ways but the final tree is insensitive to the algorithm selected. Another method is explained in the following for easy understanding. Let us assume that the complexity parameter's initial value is zero. Now for every tree (including the first, containing only the root node), compute the value for the function defined as the costs for the tree plus the complexity parameter times the tree size. Increase the complexity parameter continuously until the value of the function for the largest tree exceeds the value of the function for a smaller-sized tree to be the new largest tree, continue increasing the complexity parameter continuously until the value of the function for the largest tree exceeds the value of the function for a smaller-sized tree, and continue the process until the root node is the largest tree. Those who are familiar with numerical analysis will recognize the use of a penalty function in this algorithm. The function is a linear combination of the costs, which generally decrease with tree size, and tree size, which increases linearly. As the complexity parameter is increased, larger trees are penalized for their complexity more and more, until a discrete threshold is reached at which a smaller-sized tree's higher cost is overweighed by the largest tree's higher complexity [Hill *et al.*, 2006].

The sequence of largest trees obtained by this algorithm has a number of interesting properties. They are nested, because successively pruned trees contain all the nodes of the next smaller tree in the sequence. Initially many nodes are often pruned going from one tree to the next smaller tree in the sequence, but fewer nodes tend to be pruned as the root node is approached. The sequence of the largest trees is optimally pruned, because for every size of the tree in the sequence, there is no other tree of the same size with lower costs. Proofs and/or explanations of these properties can be found in Breiman *et al*. (1984).

Apart from visualizing the problem domain in a graphical form, the final tree can be used to make inference for new data, too. Every data point can be run down the tree using the splitting criteria and the class of the data will be the dominating class of the node where it ends up. This study uses rpart package of the R program [Dalgaard, 2008] to conduct the activities related to CART. Interested readers are requested to consult Soman *et al*. (2006) for further details.

3.6 Summary

The following figure (Figure 3.7) summarizes sequentially the work flow and methodical steps of this thesis.

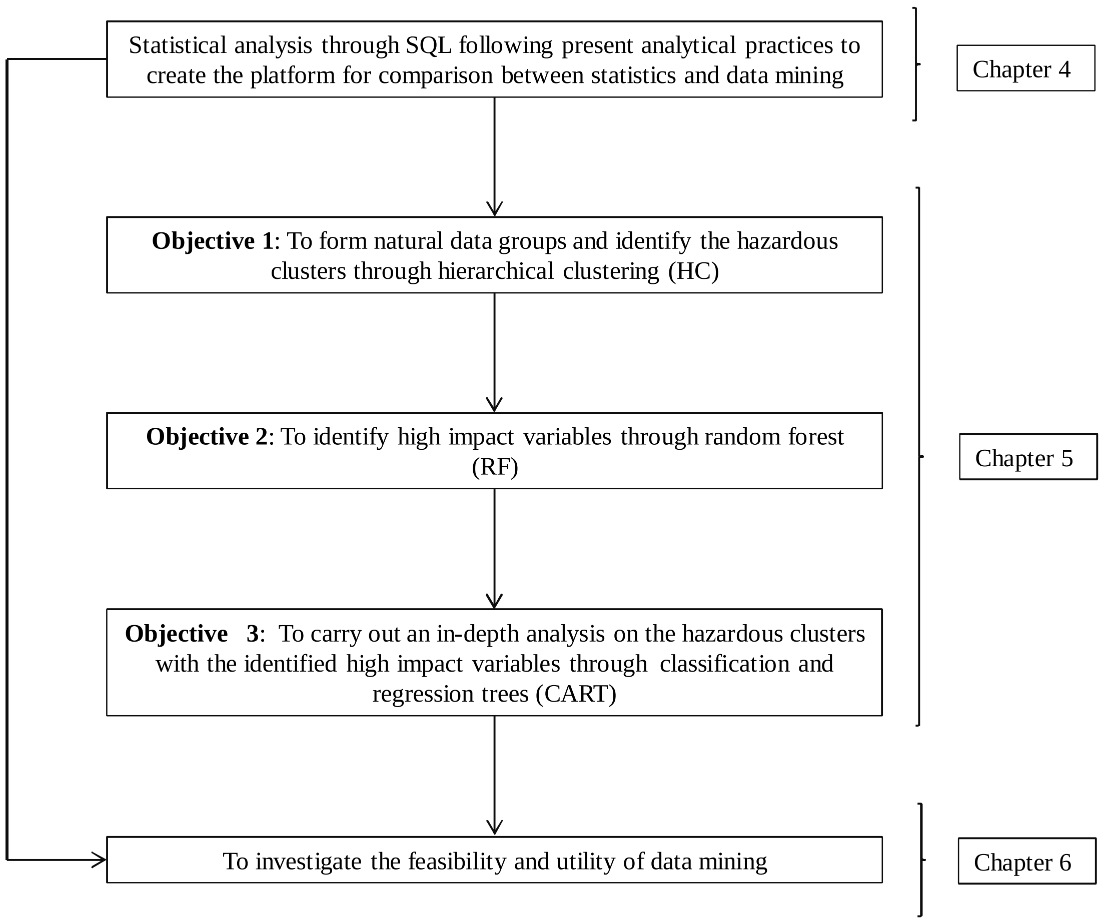


Figure 3.7. Work flow of the study.

CHAPTER 4
ROAD SAFETY IN BANGLADESH AT A GLANCE

4.1 Introduction

Although the official road accident data of Bangladesh indicated the meliorating scenario of this sector, the actual impression is just the opposite. Road accident incidences have made an enduring place in print and electronic media with other headline creating news. WHO (2009) estimates the actual fatalities from road crashes could well be 20,000 each year taking consideration of underreporting and definitional inconsistencies while in the police reported statistics it is around 3,000 each year. In economic terms, road crashes in Bangladesh are costing the community nearly 2 percent of GDP.

In Bangladesh underreporting of road accidents remains a huge problem in the country and the situation is even worse with regard to non-fatal injuries. Furthermore, improper transcription and recording of accident report forms (ARFs), lack of training and other demand at the accident scene and posterior induce internal inaccuracy in accident database. Moreover, the present ARF is inadequate to provide detail, in-depth and real scenario of crashes. Therefore, it becomes difficult to quantify the actual magnitude, trend, characteristics and identify the factors responsible through general statistical tools.

However, in this chapter an attempt has been made to depict magnitude, trends, and characteristics of prevailing road safety situation through the existing general statistical analytical practices (viz. cross tabulations, graphs etc.) in Bangladesh. Additionally, the accident database was analyzed dividing it into four major parts viz. pedestrian accident database, double vehicle accident database, single vehicle accident database and multi vehicle accident database for ease in comparison and interpretation of results with data mining outcomes as outlined in Chapter 3. The basic framework for road accident database is also explicated in brief along with its constraints for better understanding of data limitations.

4.2 Road Accident Database System in Bangladesh

In Bangladesh police is responsible for road accident data collection and storage as they are the most widespread organization and able to reach remote parts of the country. Before 1996, there was no specific format for accident data collection. At that time, information had been collected by police stations, locally known as thanas. The information were then accumulated in the form of aggregate reports and passed on to districts and metropolitan police offices on a monthly basis. Finally the data were assembled in the police headquarters (HQ) for official road accident statistics. The statistics were extremely limited in scope to be used in research or engineering purposes.

Bangladesh Police, in collaboration with Institutional Development Component (IDC), introduced a new ARF which was experimentally inaugurated into the northern division of Dhaka Metropolitan Police (DMP) area in June 1995. IDC of the Second Road Rehabilitation and Maintenance Project (RRMP2) was funded by Department for International Development (DFID) of British Government. By the end of 1996 all the police stations of DMP were brought under the network. The new scheme resulted in substantial betterment in accident information system of the country. The whole system was computerized through the application of Microcomputer Accident Analysis Package (MAAP) developed by the Transport Research Laboratory (TRL) of the United Kingdom (UK) specifically for storage and analysis of road accident data. This reporting system has been in use throughout the country since 1997 and it has become a mandatory responsibility [Regulation 254(b)] of police from September 1999.

For any type of accident, First Information Report (FIR) is filed by a sub-inspector of police. In case of road traffic accident this officer needs to complete an ARF additionally after visiting the accident spot and clarifying the information. The ARF is then dispatched to the respective Accident Data Units (ADU) where the information of ARF and location of the accident is incorporated in MAAP. Ten regional ADUs were established during early 1998. These units are responsible for processing and analysis of road accident data in their jurisdictions. Recently two more ADUs have been established but they are yet to become functional (Table 4.1).

Table 4.1. Regional ADUs and their jurisdictions

Location of ADUs	Zonal Jurisdiction
DMP	Dhaka Metropolitan Area
Dhaka Range	Dhaka Division (Except DMP Area)
CMP	Chittagong Metropolitan Area
Chittagong Range	Chittagong Division (Except CMP Area)
RMP	Rajshahi Metropolitan Area
Rajshahi Range	Rajshahi Division (Except RMP Area)
KMP	Khulna Metropolitan Area
Khulna Range	Khulna Division (Except KMP Area)
Sylhet Range	Sylhet Division
Barisal Range	Barisal Division
SMP	Yet to become functional
BMP	Yet to become functional

To assemble the national accident database and to analyze the data an additional ADU was established at the police HQ. Data are collected from the regional ADUs in soft (MAAP) format for preservation and to use as a source of intelligence.

The Accident Research Institute (ARI) of Bangladesh University of Engineering and Technology (BUET) essentially uses the MAAP database for research purposes. This database was transferred to ARI with institutional collaboration of the Road Safety Cell (RSC) of Bangladesh Road Transport Authority (BRTA) and the police department. Current road safety research and investigation works have been based on this database. However, to strengthen the database information, ARI collects the hard copies (ARFs) and soft copies (MAAP) from ADUs, add Road User Movement (RUM) codes to facilitate data analysis and modify, validate and fill up the missing information into MAAP as extracted from corrected ARFs. Bengali format of the ARF (currently in use), its English format, and the instruction guide for filling up the ARF is enclosed in Appendix A, Appendix B, and Appendix C sequentially for clear understanding of the present road accident database system in Bangladesh.

4.3 Road Safety Status and Analytical Practices in Bangladesh

In this section of the thesis an attempt has been made to present the road safety status of the country during 1998-2010 through general statistical practices. These analytical practices include generating tables, producing graphs, etc. and it is to be noted that these crude techniques have been the only analytical basis for road traffic accident analysis in the country so far. This study is concerned about how accident severities are related to road and roadway environment, and operating conditions. Therefore, predictors related to these issues have been analyzed against years through SQL to represent the magnitude, trends, and characteristics of the accidents. The outcomes are presented in the following sections according to the ARF's variables sequence. However, the source of accident data and how it is incorporated in this study is outlined in Chapter 5.

Additionally, the accident database have been analyzed dividing it into four major parts viz. pedestrian accident database, double vehicle accident database, single vehicle accident database and multi vehicle accident database for ease in comparison and interpretation of results and the generated tables are incorporated in Appendix D to Appendix G consecutively.

4.3.1 Year-wise accident severities

Accident severity analysis (Tables 4.2, 4.3, 4.4 and 4.5) revealed an interesting fact about the accident database. For all four cases (pedestrian accident, double vehicle accident, single vehicle accident and multi vehicle accident) fatal accident percentage is found to be the highest (80.39%, 54.03%, 67.55% and 41.26% chronologically). It is supposed to be in the reverse order i.e. motor collision/property damage only (PDOs) accidents should have been of the highest percentages. Except multi vehicle accidents, all other accidents are decreasing in recent years according to the database, which is quite farfetched. Thus it becomes clear that accidents with hefty consequences and a certain percentage of fatal accidents are reported and accordingly recorded in the accident database. Furthermore, it is to be noted that pedestrian accidents especially pedestrian fatal accidents are of great concerns for the country. However, these statistical tables failed to provide any further information regarding these accident events.

Table 4.2. Year-wise pedestrian accident severities

Year	Accident Severity				Total
	F	G	S	M	
2010	835	116	10	NA	961
2009	1044	174	21	NA	1239
2008	1490	257	48	NA	1795
2007	1849	288	49	NA	2186
2006	1208	193	28	NA	1429
2005	981	170	31	NA	1182
2004	1375	250	50	NA	1675
2003	1334	274	44	NA	1652
2002	1527	362	38	NA	1927
2001	1087	240	28	NA	1355
2000	1400	395	49	NA	1844
1999	1386	385	75	NA	1846
1998	1160	454	39	NA	1653
Total	16676	3558	510	NA	20744

Note: F=Fatal accident, G=Grievous accident, S=Simple injury accident, M=Motor collision/property damage only (PDO) accident, NA=Not applicable.

Table 4.3. Year-wise double vehicle accident severities

Year	Accident Severity				Total
	F	G	S	M	
2010	493	147	27	44	711
2009	626	188	33	88	935
2008	806	299	70	111	1286
2007	726	296	86	140	1248
2006	569	200	48	98	915
2005	463	210	41	96	810
2004	622	297	109	135	1163
2003	707	374	100	135	1316
2002	684	397	111	201	1393
2001	488	280	69	93	930
2000	712	501	105	158	1476
1999	656	457	174	170	1457
1998	552	536	111	159	1358
Total	8104	4182	1084	1628	14998

Note: F=Fatal accident, G=Grievous accident, S=Simple injury accident, M=Motor collision/property damage only (PDO) accident.

Table 4.4. Year-wise single vehicle accident severities

Year	Accident Severity				Total
	F	G	S	M	
2010	155	29	7	11	202
2009	225	42	8	9	284
2008	306	62	22	15	405
2007	313	91	37	25	466
2006	243	58	22	16	339
2005	246	56	20	13	335
2004	287	67	45	14	413
2003	373	114	67	19	573
2002	387	145	50	33	615
2001	315	76	27	17	435
2000	398	127	52	47	624
1999	388	133	52	45	618
1998	277	126	40	41	484
Total	3913	1126	449	305	5793

Note: F=Fatal accident, G=Grievous accident, S=Simple injury accident, M=Motor collision/property damage only (PDO) accident.

Table 4.5. Year-wise multi vehicle accident severities

Year	Accident Severity				Total
	F	G	S	M	
2010	10	2	3	2	17
2009	9	4	3	3	19
2008	5	4	0	1	10
2007	5	4	0	1	10
2006	3	1	1	0	5
2005	4	4	0	0	8
2004	17	4	3	1	25
2003	6	4	0	4	14
2002	1	0	1	4	6
2001	0	5	1	3	9
2000	7	6	3	2	18
1999	7	11	4	5	27
1998	11	21	3	3	38
Total	85	70	22	29	206

Note: F=Fatal accident, G=Grievous accident, S=Simple injury accident, M=Motor collision/property damage only (PDO) accident.

4.3.2 Year-wise accidents by day of week, month of year and time of occurrence

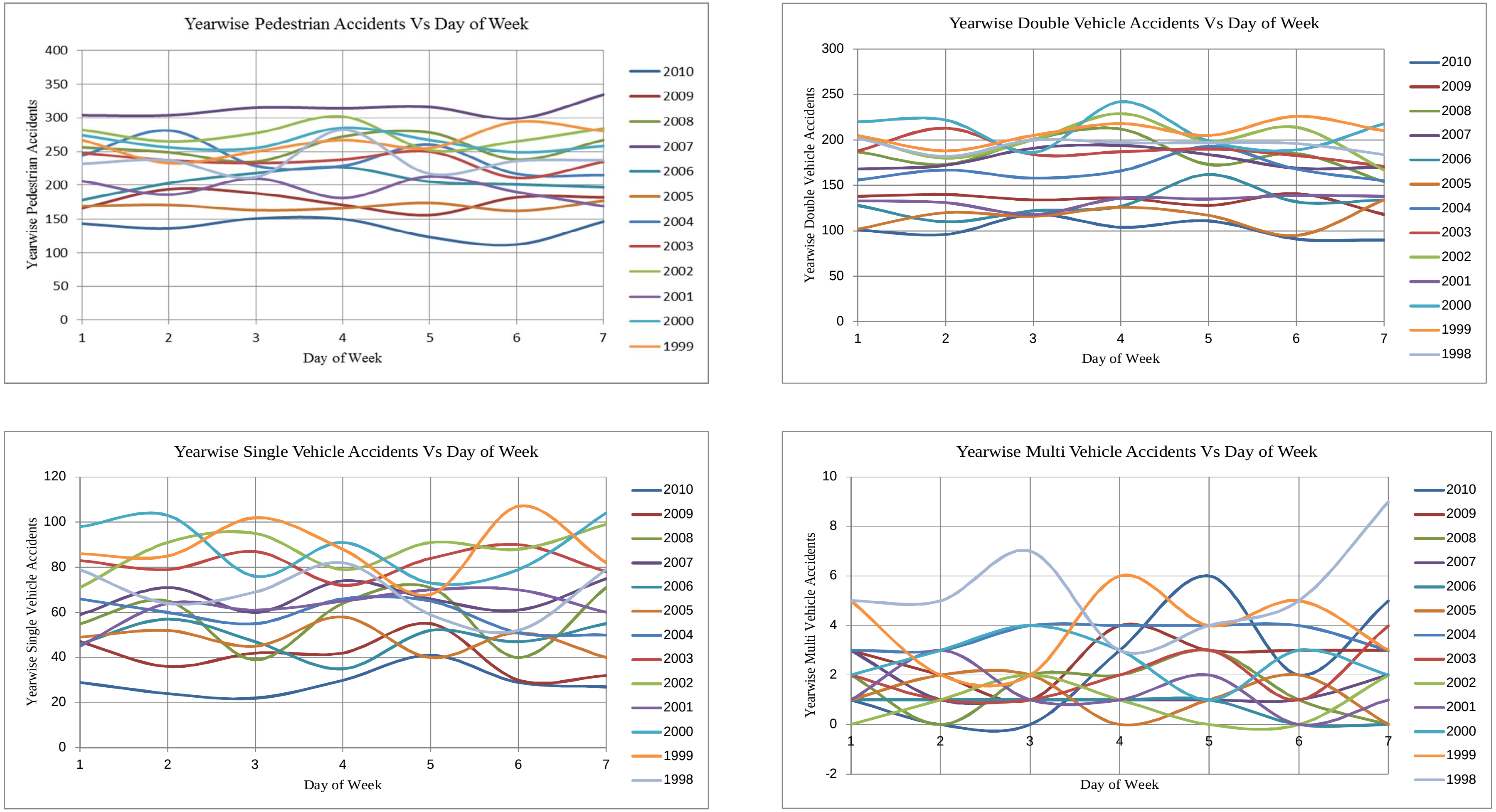
Analyses of accidents with respect to different temporal variables have been presented in Figures 4.1, 4.2 and 4.3. No significant trend of accidents has been identified with respect to day of week and month of year. However, it is perceived that when number of samples comes down the fluctuation increases. On the other hand, accident analysis with respect to time of occurrence has identified 10 am to 1 pm and 3 pm to 6 pm have been the most crucial hours of accident events for pedestrian accidents, double vehicle accidents and single vehicle accidents; yet multi vehicle accidents did not provide any noteworthy scenario.

4.3.3 Year-wise accidents by junction type

Mid-block sections of roads are more accident prone than junctions as depicted by Figure 4.4 and this is valid for all four categories of accidents i.e. pedestrian accidents, double vehicle accidents, single vehicle accidents and multi vehicle accidents. More than 62 percent of these accidents have taken place at not junction sections. Other junction type has been identified as the second most susceptible segments for accidents. But it might be due to the fact that the concerned personnel were unable to fill the information correctly. Tee junctions have been prioritized as the third junction in this sequence and the recent trend is on rising side for this type.

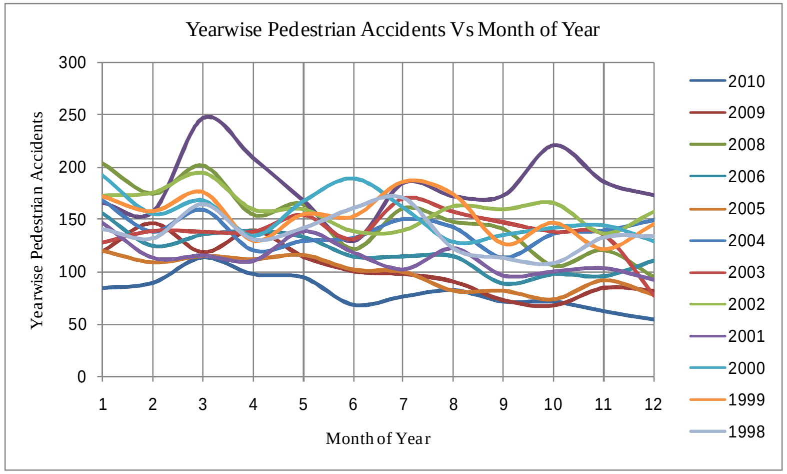
4.3.4 Year-wise accidents by traffic control system and collision type

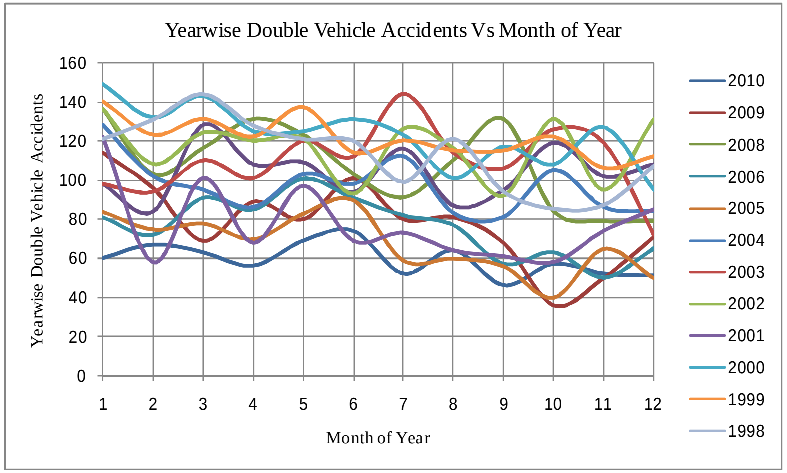
More than 72 percent of accidents have been clustered in places where there is no traffic control system available followed by other type and police controlled traffic control system (Figure 4.5). Even police controlled along with traffic light type traffic control system is also found quite ineffective in reducing accidents. On the other hand, collision type analysis identified different types of collision along with hit pedestrian accidents (90.5%) are contributing in pedestrian fatalities (Figure 4.6). For double vehicle accidents, rear end and head on; for single vehicle accidents overturn, other type and hit object off road; for multi vehicle accidents rear end and side swipe types of collisions have been found as dominant types (Figure 4.6).

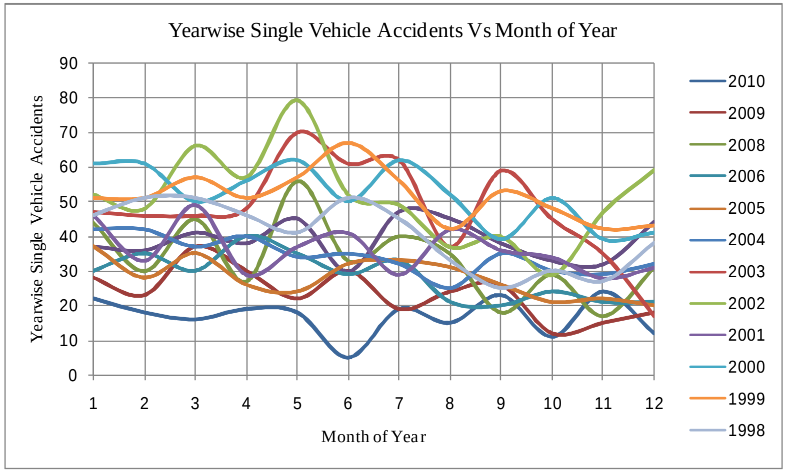


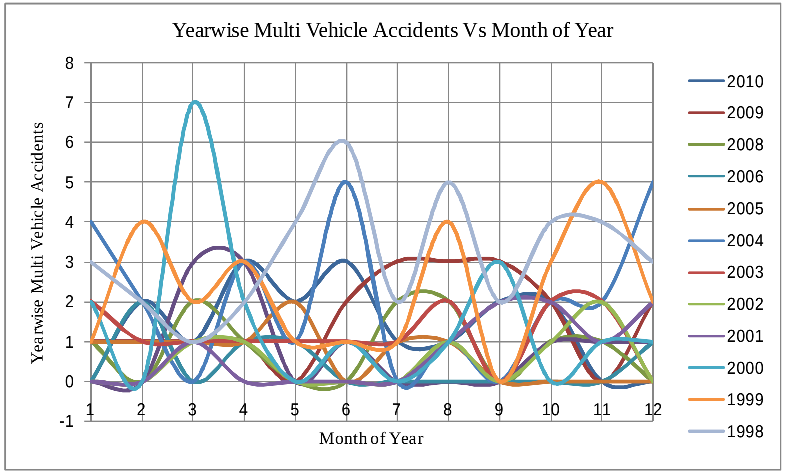
Note: 1= Monday, 2=Tuesday, 3=Wednesday, 4=Thursday, 5=Friday, 6=Saturday, 7=Sunday

Figure 4.1. Year-wise accidents vs day of week.



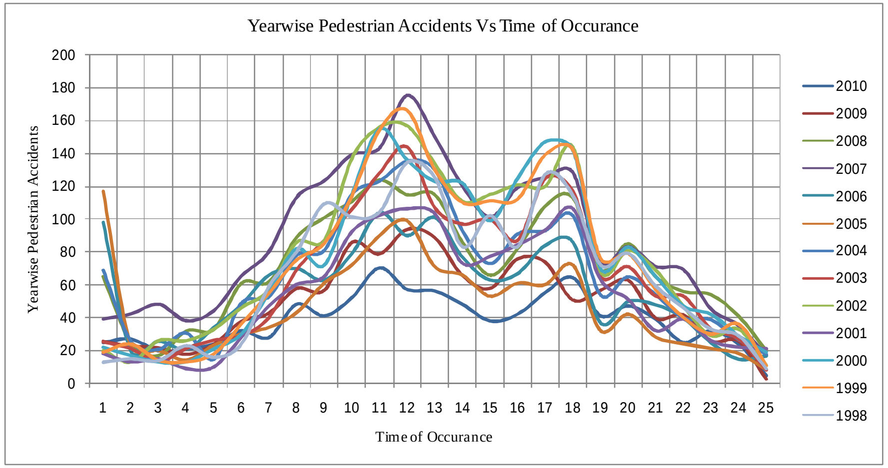


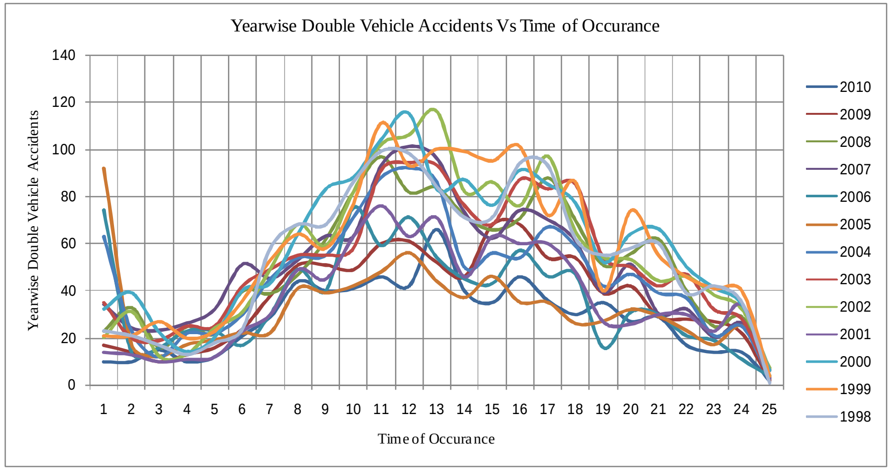


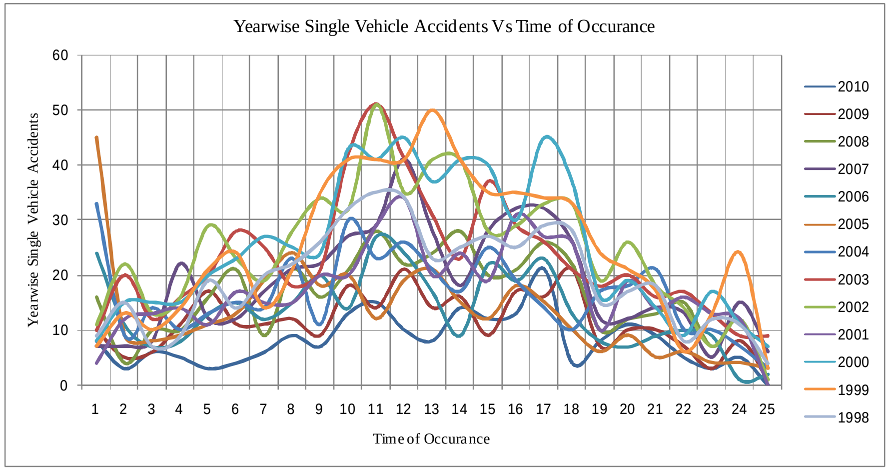


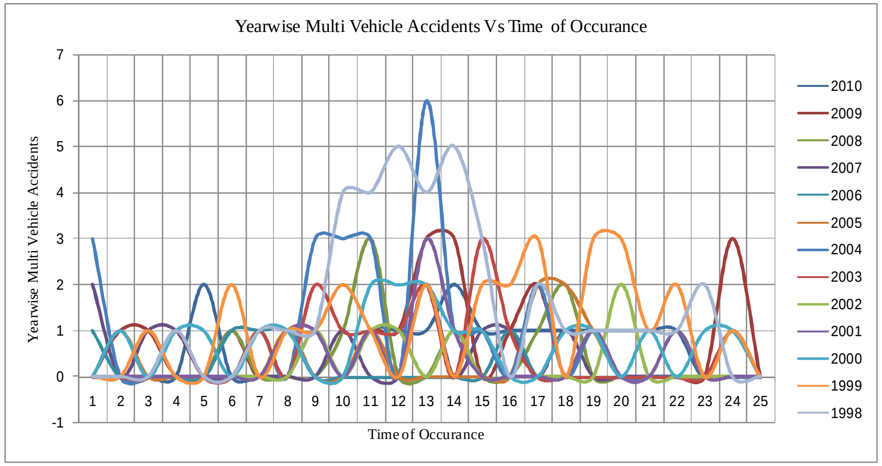
Note: 1=January, 2=February, 3=March, 4=April, 5=May, 6=June, 7=July, 8=August, 9=September, 10=October, 11=November, 12=December

Figure 4.2. Year-wise accidents vs month of year.



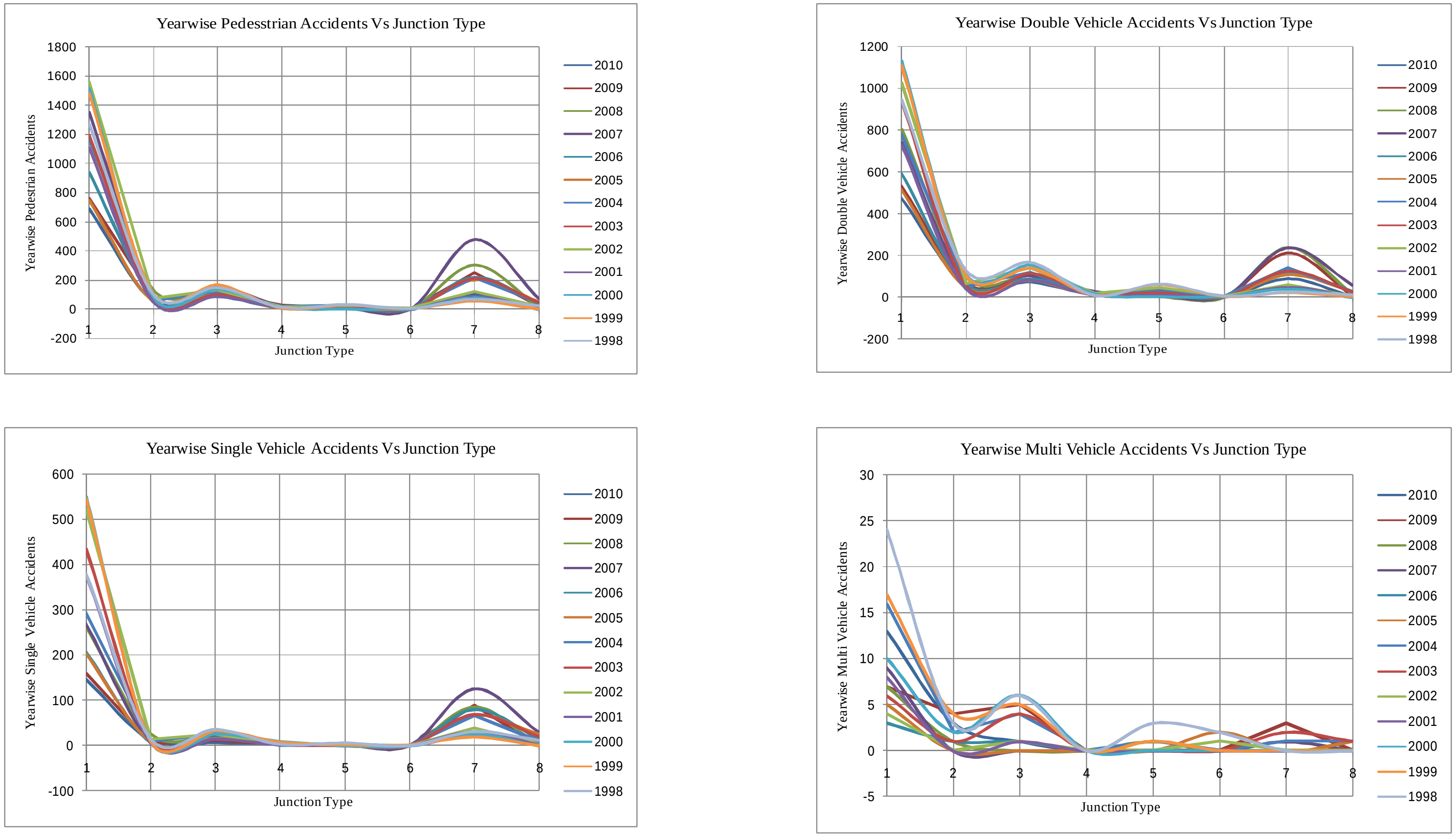






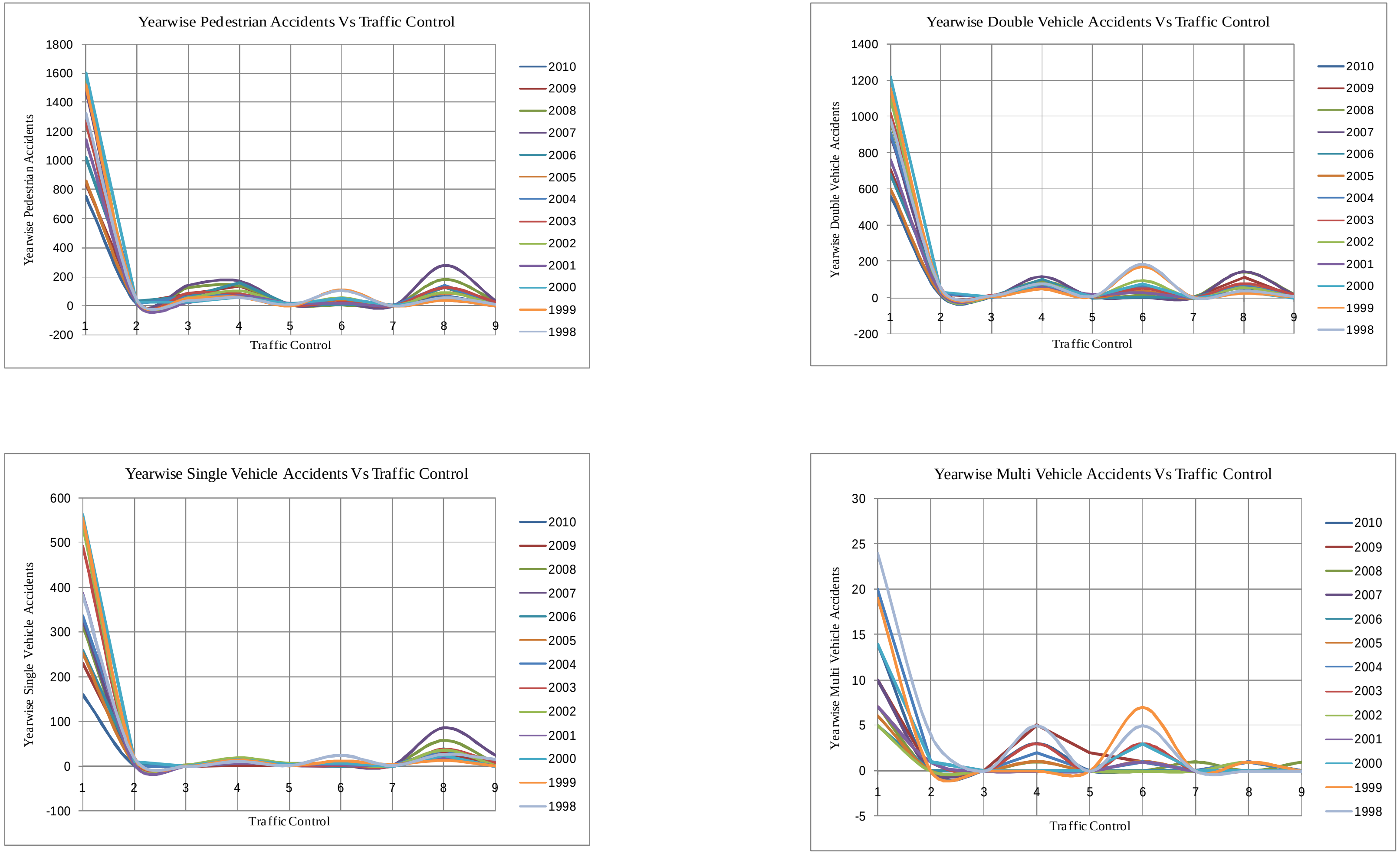
Note: 25=?/Blank data field

Figure 4.3. Year-wise accidents vs time of occurrence.



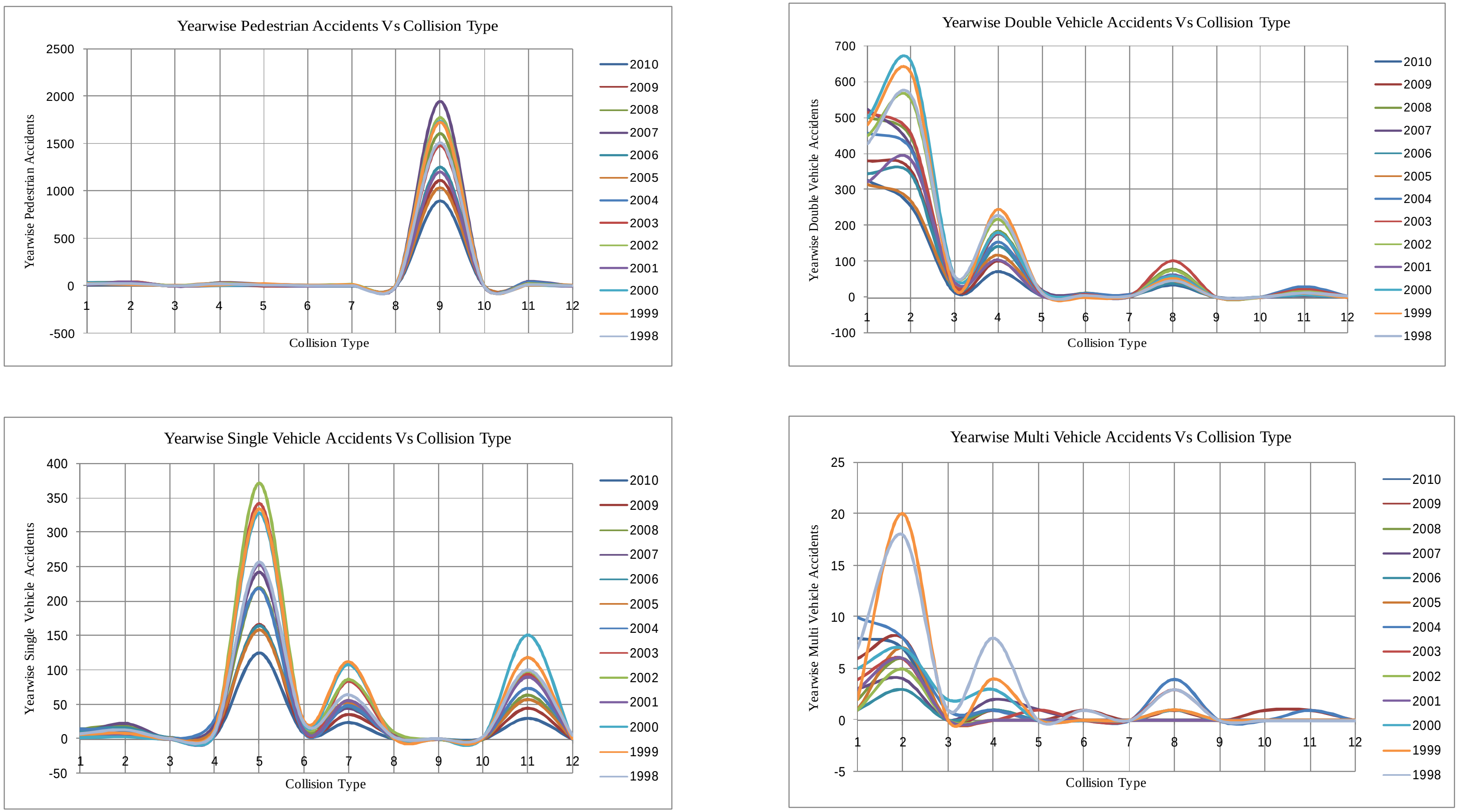
Note: 1=Not at junction, 2=Cross junction, 3=Tee junction, 4=Staggered junction, 5=Roundabout, 6=Railway/Level crossing, 7=Other, 8=?/Blank data field

Figure 4.4. Year-wise accidents vs junction type.



Note: 1=No control, 2=Centerline marking, 3=Pedestrian crossing, 4=Police controlled, 5=Traffic lights, 6=Police+Traffic lights, 7=Stop/Give way sign, 8=Other, 9=?/Blank

Figure 4.5. Year-wise accidents vs traffic control system.



Note: 1=Head on, 2=Rear end, 3=Right angle, 4=Side swipe, 5=Overturn, 6=Hit object in road, 7=Hit object off road, 8=Hit parked vehicle, 9=Hit pedestrian, 10=Hit animal, 11=Other, 12=?/Blank data field

Figure 4.6. Year-wise accidents vs collision type.

4.3.5 Accidents by traffic movement and presence of road dividers

Database revealed that more than 80 percent of road traffic accidents are occurring in two-way (movement) roads (Figure 4.7); and nearly 75 percent of these accidents have taken place in roads without dividers (Figure 4.8). Year-wise detail distribution of these accidents (pedestrian accidents, double vehicle accidents, single vehicle accidents, and multi vehicle accidents) is presented in Appendix D to Appendix G in Tables 7 and 8.

4.3.6 Accidents by weather and light condition

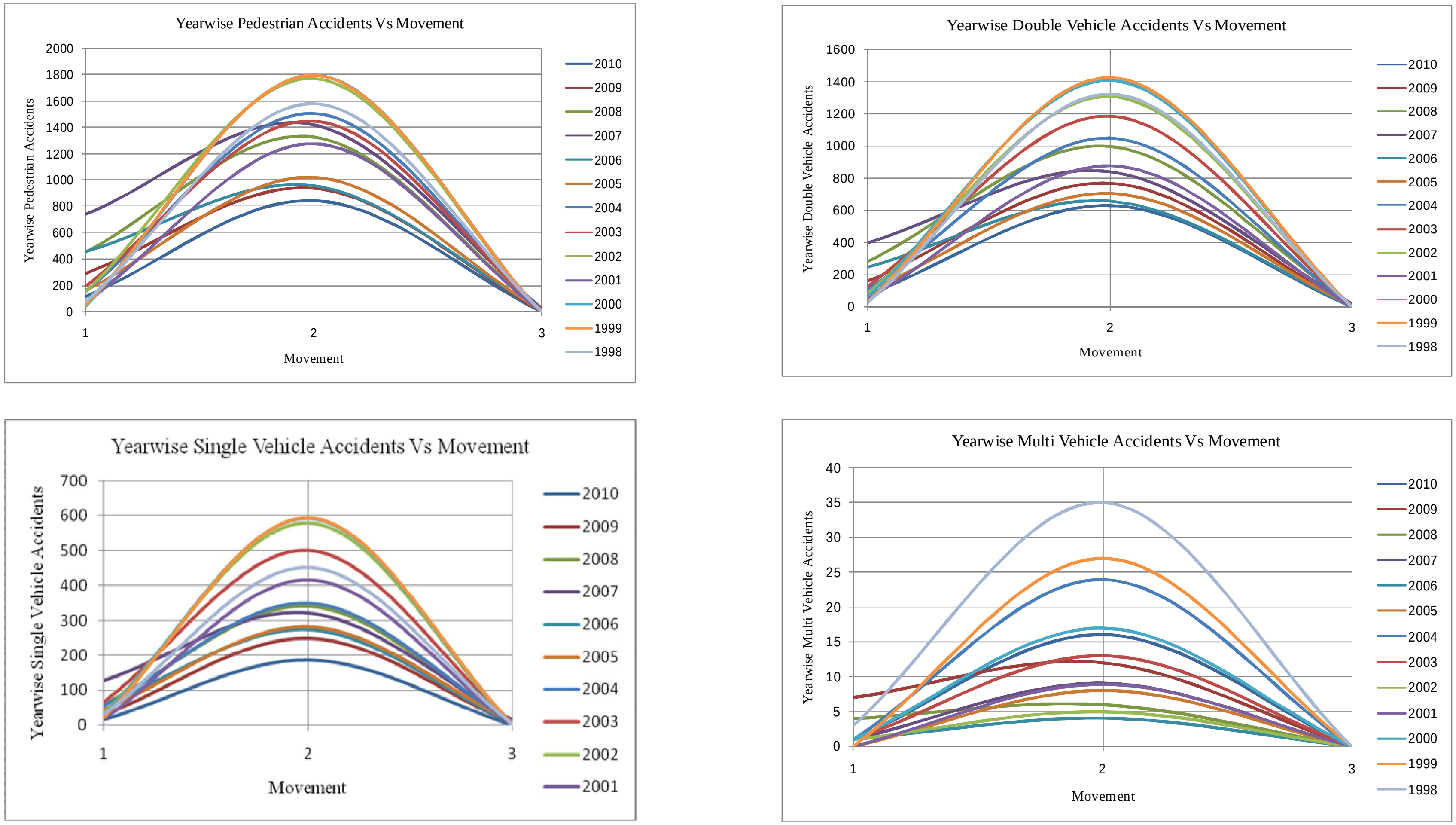
Fair weather and daylight have been identified as stimulating factors for all types of accidents (Figures 4.9 and 4.10). General statistics have been effective here only to find out the percentages of crash occurrence in different meteorological conditions but failed to give an insight into the actual scenarios. This highlights the scope limitations in normal charting techniques and graph generations which is able to elicit the general trends only.

4.3.7 Accidents by road geometry, and surface condition, type and quality

Analyses of pedestrian accidents, double vehicle accidents, and multi vehicle accidents unveiled that more than 90 percent of these accidents are affiliated with straight and flat road geometry, and dry, sealed and good road surface conditions (Figures 4.11, 4.12, 4.13 and 4.14). For single vehicle accidents the same is true for more than 82 percent cases (Figures 4.11–4.14). Year-wise detail trends of these predictors for accident occurrence are encompassed in Appendix D to Appendix G in Tables 11, 12, 13 and 14.

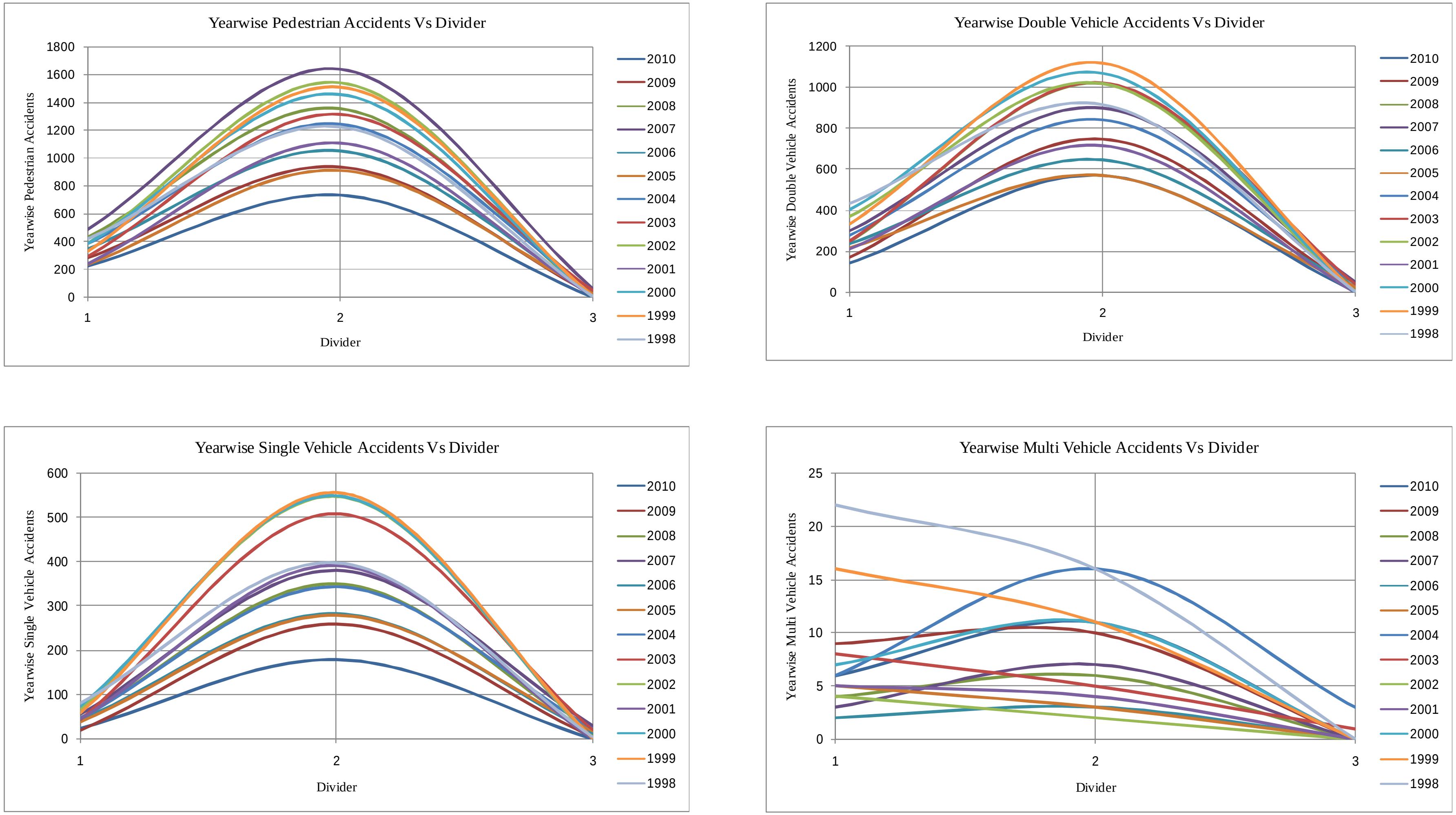
4.3.8 Accidents by road class, road feature, and location

In case of pedestrian accidents, national highways (38.54%) and city roads (25.35%) have been spotted with highest percentages of accidents (Figure 4.15). In addition, these accidents are associated with normal road features (96.17% cases, Figure 4.16) and distributed quite similarly (Figure 4.17) in rural (nearly 60%, decreasing trend) and urban areas (nearly 40%, increasing trend). Detail trends are incorporated in Appendix D in Tables 15, 16 and 17.



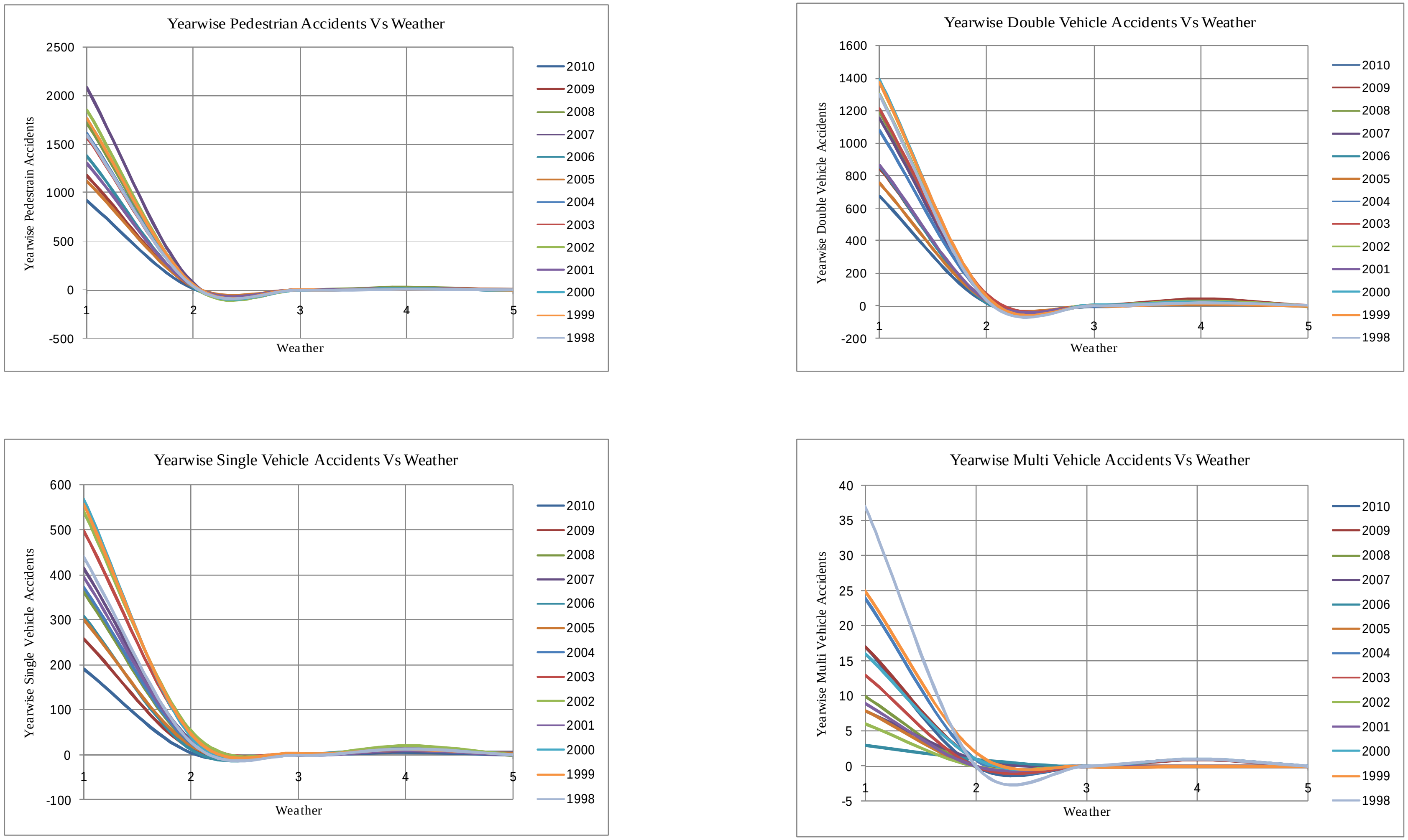
Note: 1=One-way movement, 2=Two-way movement, 3=?/Blank data field

Figure 4.7. Year-wise accidents vs traffic movement.



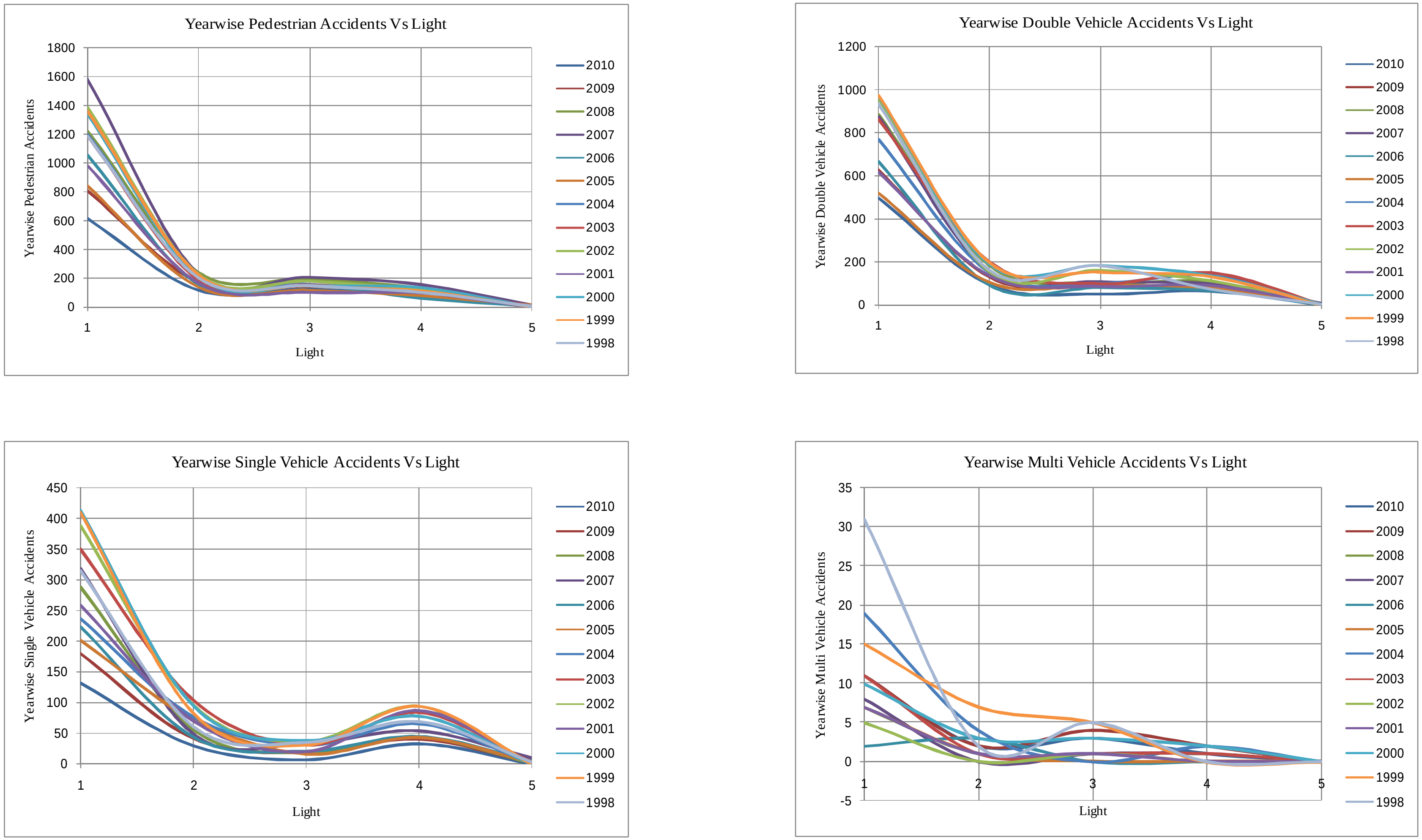
Note: 1=Yes, 2=No, 3=?/Blank data field

Figure 4.8. Year-wise accidents vs presence of divider in roads.



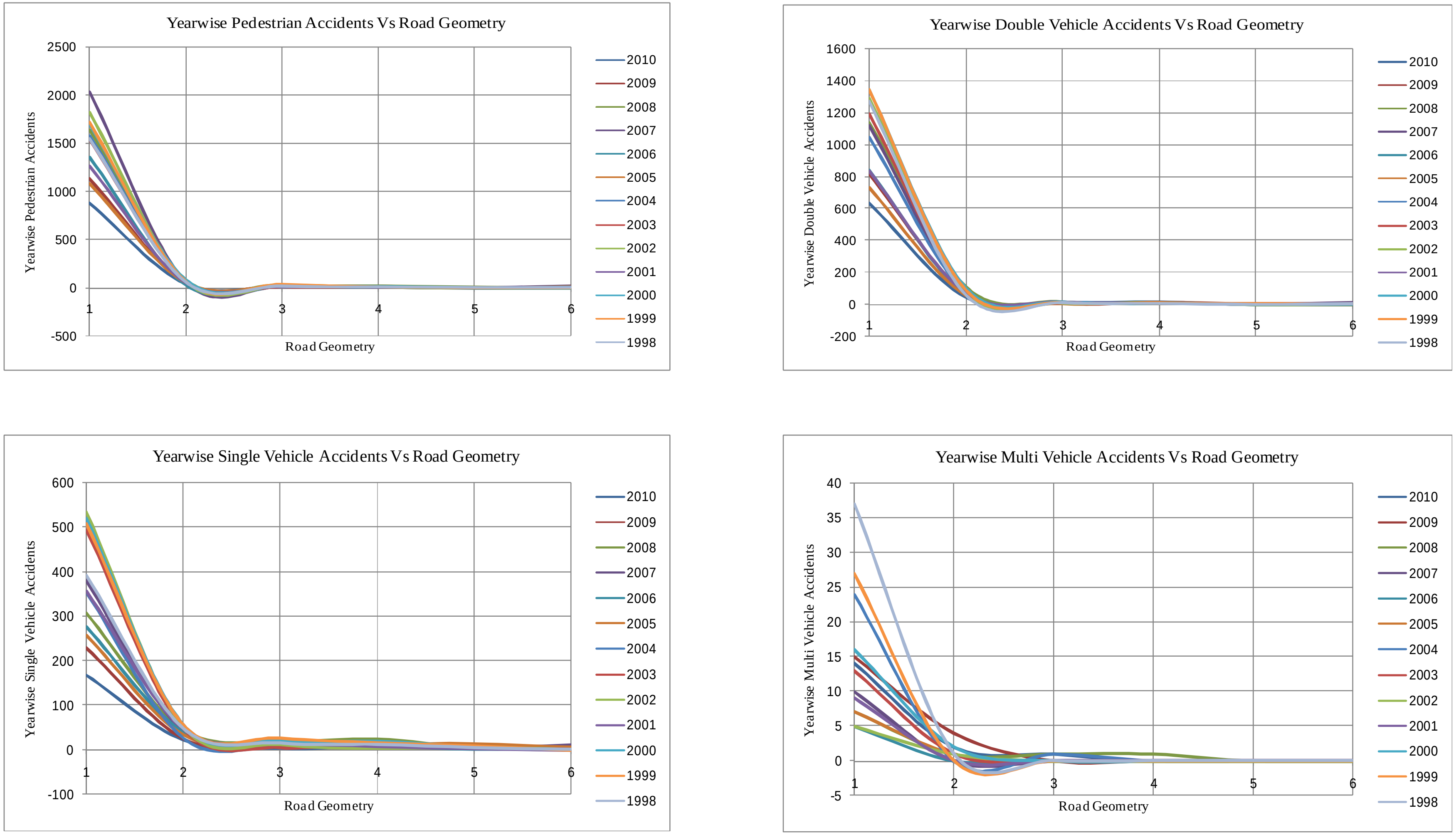
Note: 1=Fair, 2=Rain, 3=Wind, 4=Fog, 5=?/Blank data field

Figure 4.9. Year-wise accidents vs weather condition.



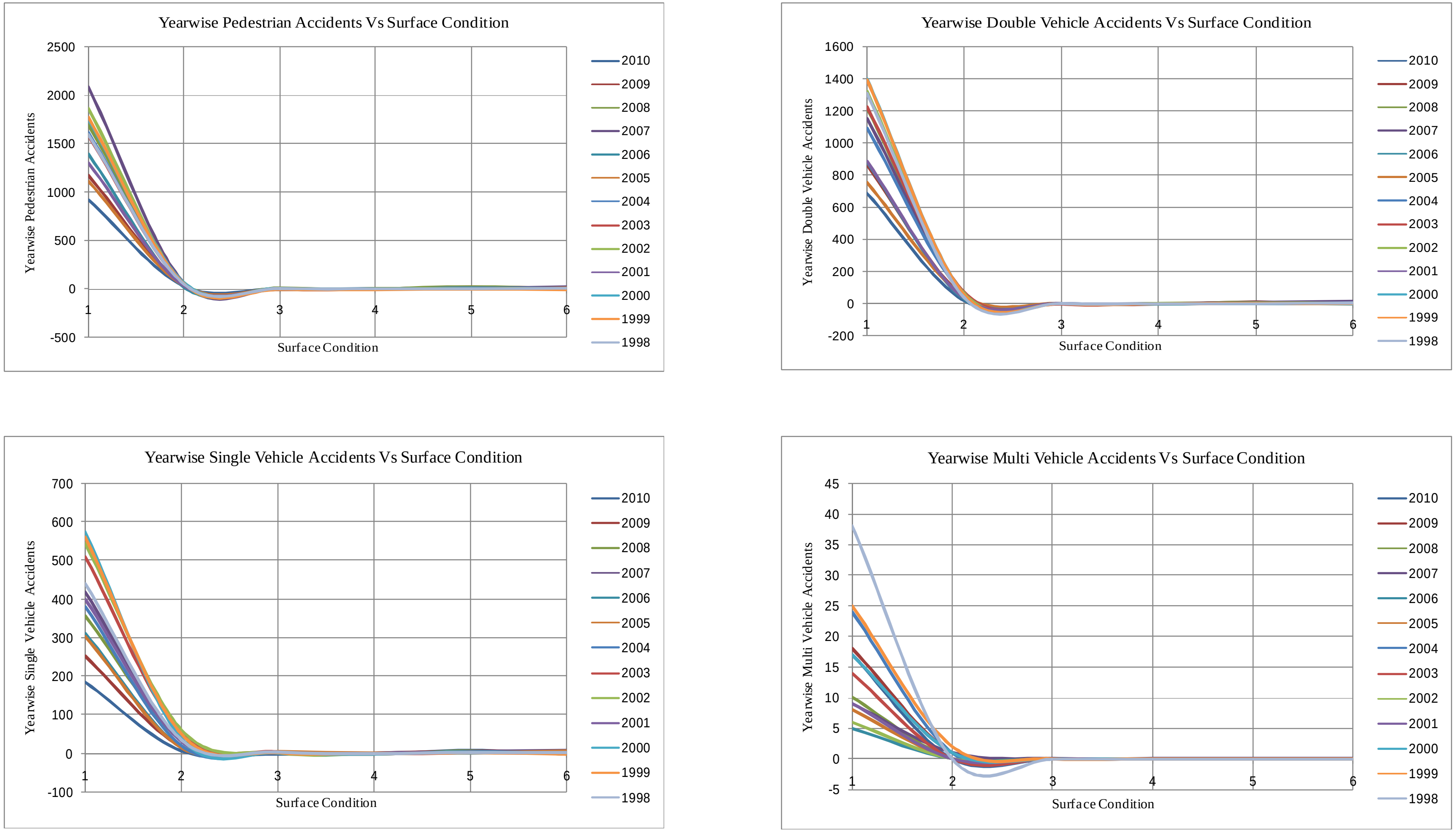
Note: 1=Daylight, 2=Dawn/Dusk, 3=Night (lit), 4=Night (unlit), 5=?/Blank data field

Figure 4.10. Year-wise accidents vs light condition.



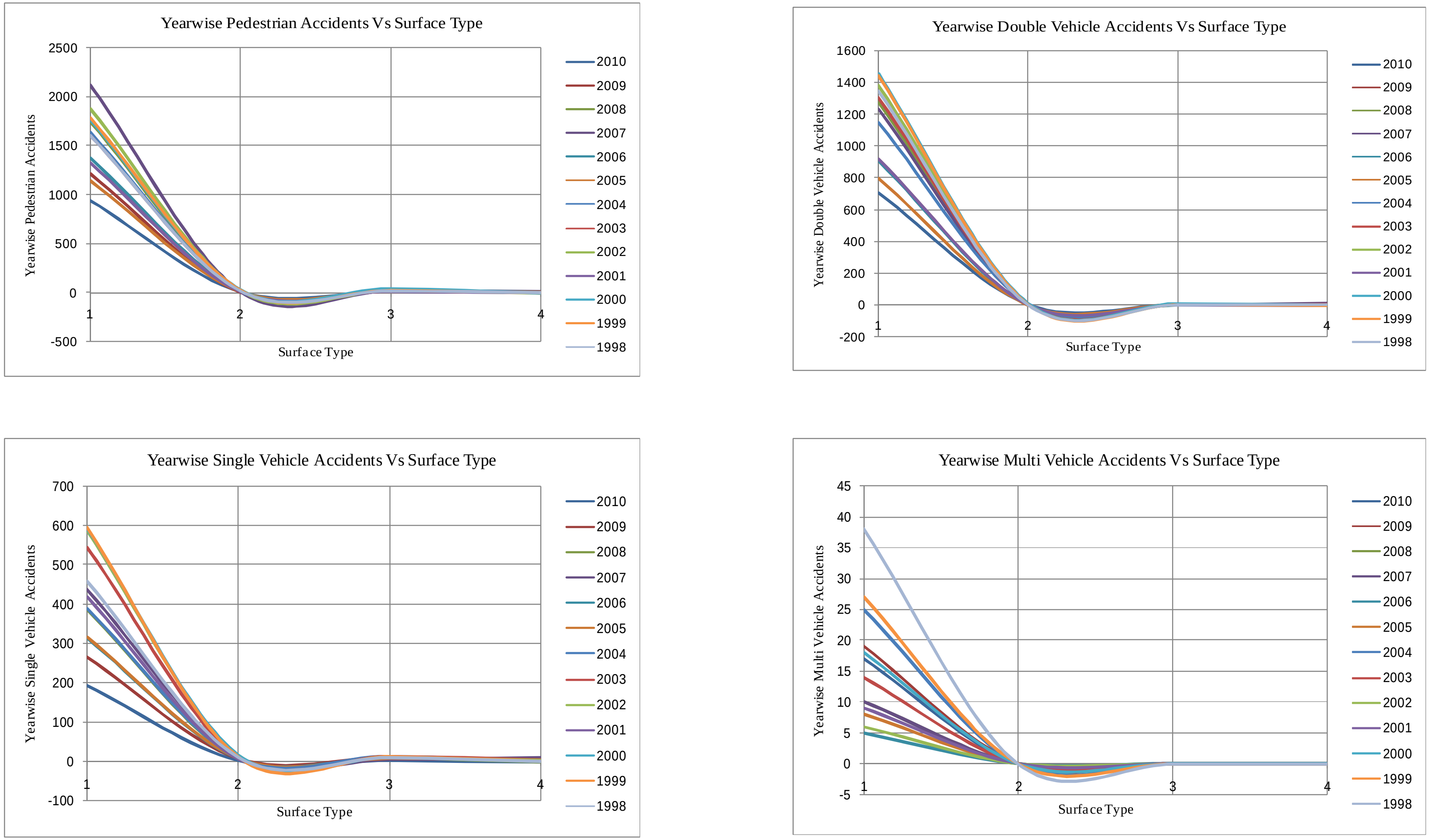
Note: 1=Straight+Flat, 2=Curve only, 3=Slope only, 4=Curve+Slope, 5=Crest, 6=?/Blank data field

Figure 4.11. Year-wise accidents vs geometric condition of road.



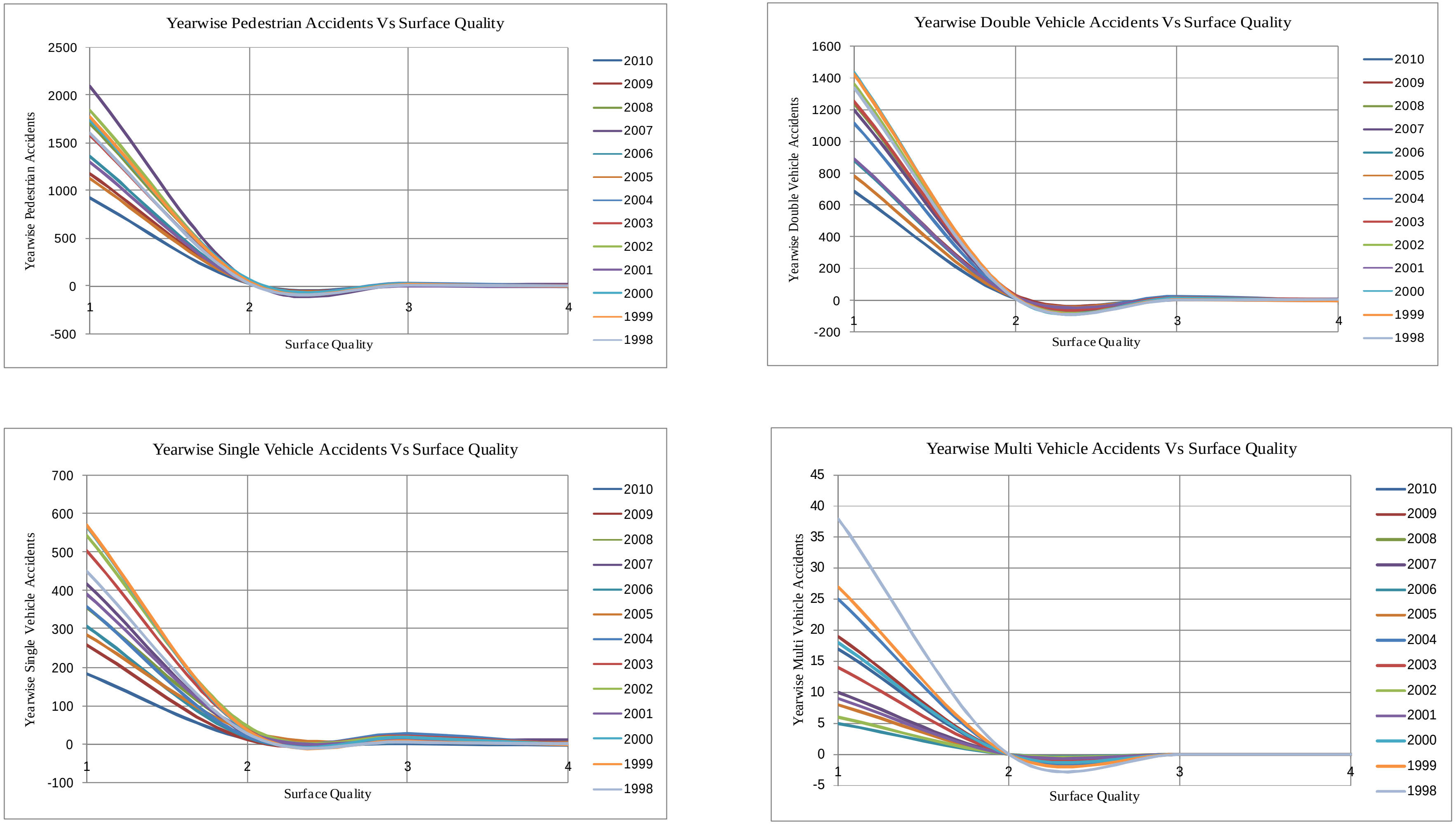
Note: 1=Dry, 2=Wet, 3=Muddy, 4=Flooded, 5=Other, 6=?/Blank data field

Figure 4.12. Year-wise accidents vs road surface condition.



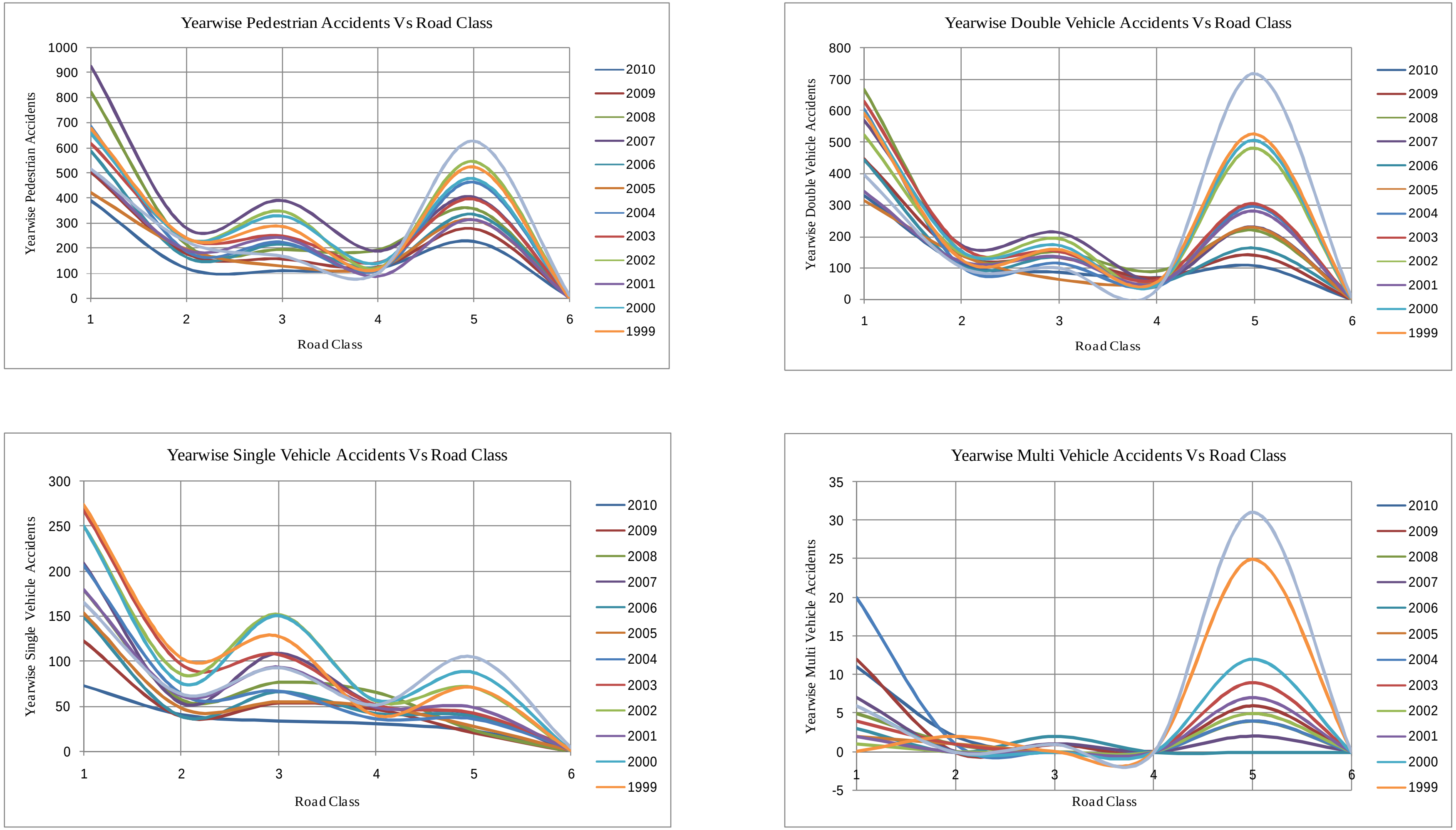
Note: 1=Sealed, 2=Brick, 3=Earth, 4=?/Blank data field

Figure 4.13. Year-wise accidents vs road surface type.



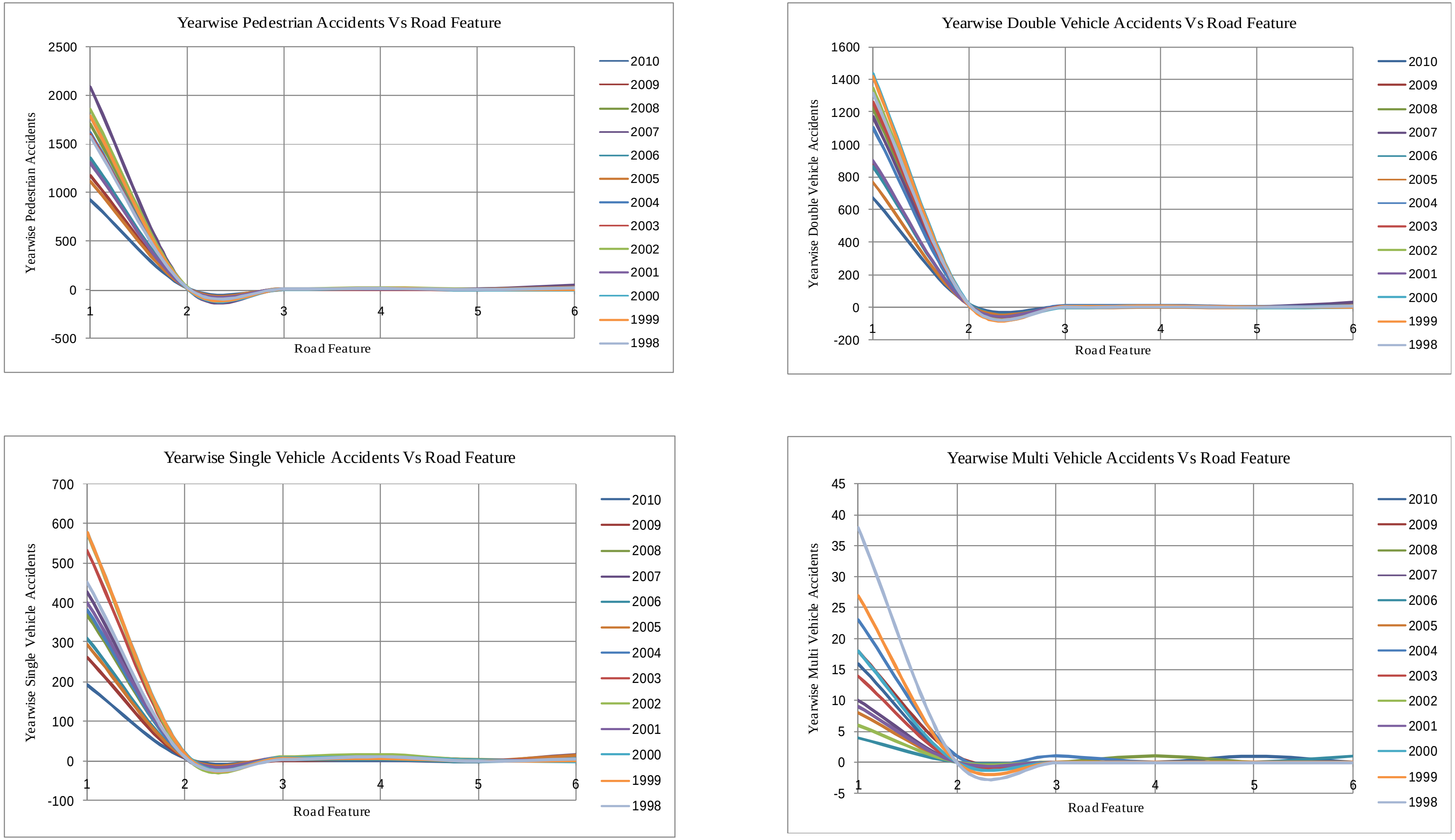
Note: 1=Good, 2=Rough, 3=Under repair, 4=?/Blank data field

Figure 4.14. Year-wise accidents vs surface quality of road.



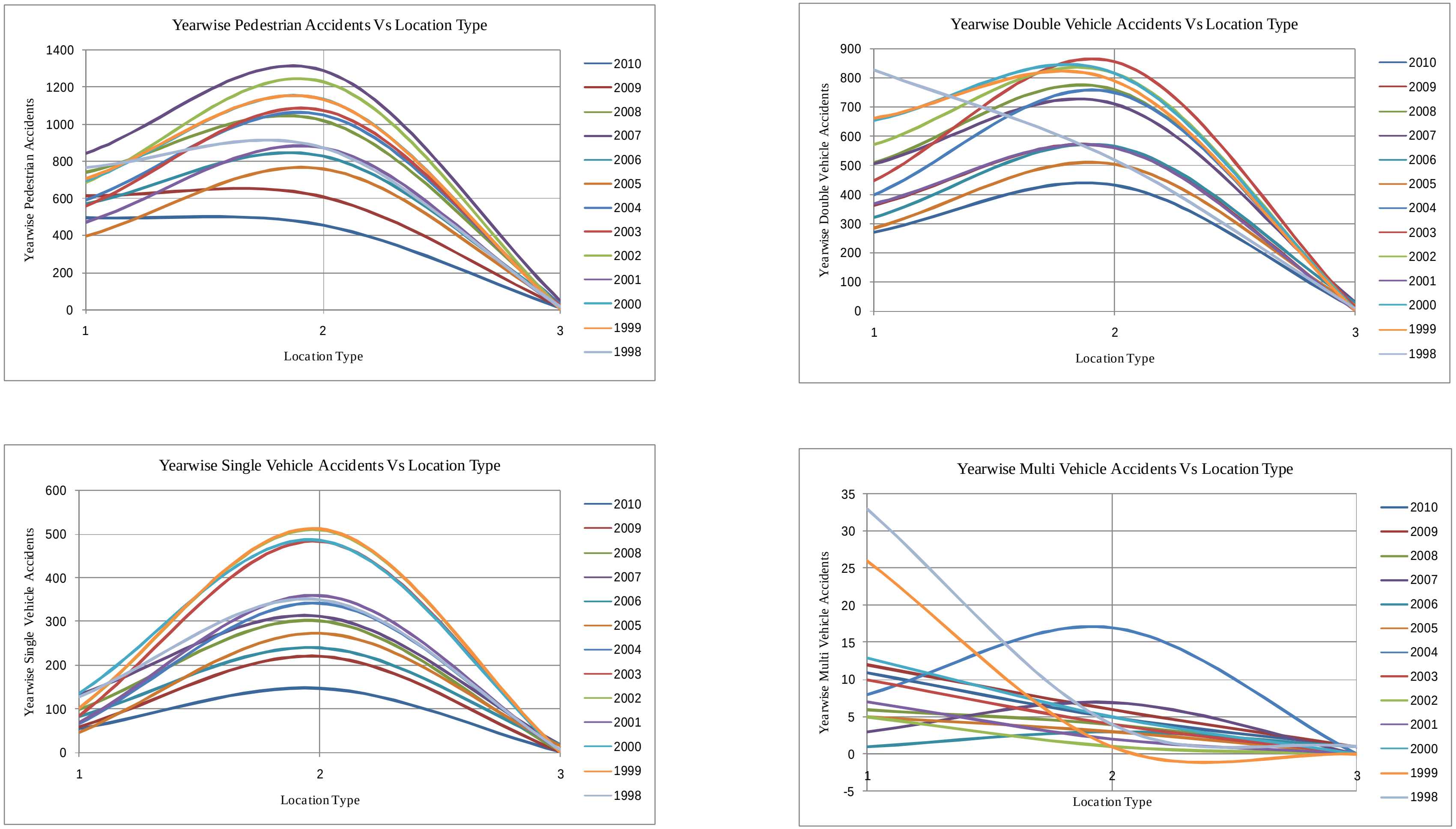
Note: 1=National highway, 2=Regional highway, 3=Feeder road, 4=Rural road, 5=City road, 6=?/Blank data field

Figure 4.15. Year-wise accidents vs road class.



Note: 1=None, 2=Bridge, 3=Culvert, 4=Narrowing/Restriction, 5=Speed breakers, 6=?/Blank data field

Figure 4.16. Year-wise accidents vs road feature.



Note: 1=Urban area, 2=Rural area, 3=?/Blank data field

Figure 4.17. Year-wise accidents vs location.

Double vehicle accidents have also been found huddled in national highways (43.08%) and city roads (28.04%). Furthermore, these accidents are affiliated with normal roadway features (96.15%) and are distributed quite similarly in rural (57.61%, decreasing trend) and urban areas (41.25%, increasing trend). Yearly details are unified in Figures 4.15, 4.16 and 4.17 and in Appendix E in Tables 15, 16 and 17.

On the other hand, feeder roads (20.51% accidents) have been identified as vulnerable carriageways along with national highways (42.86% accidents) for single vehicle accidents (Figure 4.15). Again, it has been underscored that this type of accidents are more prone to rural areas (78.34% accidents; Appendix F, Table 17). Multi vehicle accidents have followed the same trends (in case of road class and road feature) as pedestrian accidents and double vehicle accidents. But they are more prone to urban areas (67.96%, Figures 4.15, 4.16, 4.17 and Tables 15, 16 and 17 in Appendix G) than rural areas.

From the statistical analyses presented above, a gross idea regarding the current analytical techniques of road traffic accidents along with a general status about the present state of the problem have been generated. This has served as the platform/basis for comparison between statistical outcomes and data mining findings for this thesis.

4.4 Constraints in Present Accident Database System

The current road traffic accident database system of Bangladesh has a number of limitations. These drawbacks can be broadly classified into three main categories – constraints in accident reporting and recording system, weaknesses of the ARF, and limitations of MAAP software regarding in-depth analyses. A brief description of these issues has been discoursed here for clear understanding of data limitations.

Constraints in accident reporting and recording: this issue has several spectra, viz. underreporting, reported but not accumulated in the database, wrong transcription and interpretation, improper recording of ARFs, lack of proper training, etc.

Underreporting – it has been perceived that if somebody does not inform police regarding a particular accident, it is quite unexpected for the police to take steps willfully to report and record an accident event. And this is particularly true for non-fatal injuries; even sometimes police tries to avoid recording a road traffic accident case or provides the counseling service in negotiation or does not know about the importance of reporting and recording of an accident event. Therefore, only the fatal cases whose consequences cannot be avoided are reported and recorded. However, it is not ensured that all the fatal accidents are properly reported even; especially the non-fatal injuries of a fatal accident are the most neglected portion in reporting.

Reported but not accumulated in the database – in case of any bad incidence police files a First Information Report (FIR); and for road accident cases they need to fill up an ARF which is expected to be recorded in MAAP later. A comparison between FIR and MAAP shows that even all the FIRs concerning road accidents are not accumulated in MAAP (Table 4.6) and may be the lower severities are being ignored in most cases.

Table 4.6. Comparison of FIR and MAAP

Year	No. of Accidents			No. of Fatalities		No. of Injuries		Total Casualties		
	FIR	MAAP	Variation (%)	FIR	MAAP	FIR	MAAP	FIR	MAAP	Variation (%)
2011	2667	NA	NA	2467	NA	1641	NA	4108	NA	NA
2010	2827	2437	14	2646	2443	1803	1706	4449	4149	7
2009	3381	2815	17	2958	2703	2686	1746	5644	4449	21
2008	4426	3800	14	3764	3570	3284	2416	7048	5986	15
2007	4869	3954	19	3749	3341	3273	2431	7022	5772	18
2006	3794	3566	6	3193	3250	2409	2412	5602	5662	-1
2005	3955	3322	16	3187	2960	2755	2570	5942	5530	7
2004	3917	3566	9	2968	3150	2752	3026	5720	6176	-8
2003	4749	4114	13	3289	3334	3818	3740	7107	7074	0
2002	4918	3941	20	3398	3053	3772	3285	7170	6338	12
2001	4091	2925	29	3109	2388	3127	2565	6236	4953	21
2000	4357	3970	9	3430	3058	1911	3485	5341	6543	-23
1999	4916	3948	20	3314	2893	3453	3469	6767	6362	6
1998	4769	3533	26	3085	2358	3997	3297	7082	5655	20

Notes: NA= Not Available, Variation % = {(FIR-MAAP)×100} ÷ FIR

Wrong transcription and interpretation – even in case of recorded accident, erroneous information can evolve from improper transcription and interpretation of ARFs. A comparison between controlled transcription of ARF at ARI and MAAP data reveals that there exits significant differences between the two. Double-entry of same information is found quite common [Alam *et al*., 2006].

Improper recording of ARF – caused due to lack of time and resources. ARFs are filled improperly thereby causing erroneous data. Many variables are not filled at all viz. location, mileage of the roadway which creates misperception [Alam *et al*., 2006].

Lack of proper training – lack of training of police officers makes it difficult for them to properly record an accident in ARF. The form is not plug and play type; without proper training it is quite difficult to fill up correctly.

Weaknesses of ARF: the present ARF is not up to date. The information recorded in this form can provide only an abstract idea about an accident but cannot able to pin point the actual scenario. For example, only 17 reasons of an accident occurrence are provided in the form that can be incorporated correctly; but these reasons are not adequate to elucidate the actual event. Even these 17 reasons will not provide the micro-level information that is required to produce a concrete conclusion. An accident may happen due to vehicle defect or may be due to tyre bursting, but the form does not provide the place to include what kind of vehicle defect it was or which tyre it was. Again the inventory that is in use to identify the accident locations is based on 1998’s status, which fails to provide present aspects needed for the analysis. There are a number of these types of shortcomings which necessitates making this form restructured.

Limitations of MAAP software: the MAAP5 software that is in use can produce cross tabulations with 2 fields of information at best. In some cases a couple of conditions may be added. This means it is possible to get ${}^{67}C_2$ number of cross tabulations (as ARF can lodge 67 fields of information so is MAAP5), but MAAP5 cannot accommodate more than 2 variables at a time. Therefore, even if all correct information is incorporated in MAAP5, it is not possible to get the best possible outcome.

CHAPTER 5
DATA MINING OF ACCIDENT DATABASE

5.1 Introduction

This chapter starts with the source of accident data and how it was incorporated for data mining applications in road traffic accident (RTA) analysis. Then the data mining applications are organized broadly into three sections. The first section explains the components of HC, which is to form natural data groups and to identify the hazardous clusters. The second section i.e. RF identifies the high impact variables and discusses the steps of choosing the proper variables for this study. Third section presents the CART analysis, results and explanation of the findings.

5.2 Data Collection

The RTA data for the period of 1998-2010 were collected from the Accident Research Institute (ARI), BUET. ARI uses Micro-computer Accident Analysis Package *five* (MAAP5) software for accident data storage and analysis purpose. The data format of MAAP5 is not compatible with R, which is the primary software for this research. Therefore, a conversion was required. M. D. Alam, ex-assistant programmer of ARI developed a tool that was able to convert the MAAP5 database to MS Excel. This tool was used for the required transformation. However, it is found that the conversion tool can not convert the whole database to Excel. Some accident records were found as garbage in the converted database and this occurrence was found as random events. Out of total 45,891 accident records, 41,741 were transferred properly; i.e. 91 percent data were successfully converted, which is adequate for the application of data mining.

5.3 Data Preparation

The Excel database extracted from MAAP5 database needed some further processing to be used for data mining. At first, the Excel format was converted to CSV (comma delimited) format; so that it can be imported by the R software. Later, it was found that the computer that was designated for the data mining operations could not handle this huge database. Therefore, it became urgent to reduce the size of this accident database.

Then a two phase approach was adopted. In the first phase, the total database was divided into four major divisions. These are pedestrian accident database (all pedestrian related accidents were brought under this division), double vehicle accident database, single vehicle accident database and multi vehicle accident database. Even after this split, the available computers could not process the required dissimilarity (distance) matrix of pedestrian accident database for hierarchical clustering.

So in the second phase, the study period was reduced to 2006-2010; ARI's last 5 years modified database. Even after these two differentiations, the four databases (pedestrian 7,610; double vehicle 5,095; single vehicle 1,696; and multi vehicle 61) were sufficient for data mining applications. However, for the general statistical analysis through SQL at MS Access (Chapter 4) the whole converted Excel database had been used.

The databases even after so many alterations were not smooth. It was difficult to pick a column of same accident variables after the roads and roadway environment and their operating condition variable related columns. It was required to modify the database manually or to develop a new tool for MAAP5 database conversion. As these two tasks were beyond this study's limit, therefore the research scope was limited to how the severities of RTAs are related to roads and roadway environment and their operating conditions.

However, the research approach applied in this thesis can be used as a framework for any type of data mining studies with high configuration computers for any size of databases.

5.4 Application of HC

The accident database was divided into four major parts viz. pedestrian accident database, double vehicle accident database, single vehicle accident database and multi vehicle accident database for ease in analysis and interpretation of results as outlined in data preparation section. HC was applied to each of these four databases separately. The cluster package of R program was used for this purpose. To reduce the complexity of the study, the algorithm was set in such a way that it evolved four dendrograms for each of the databases i.e. in total sixteen dendrograms were developed. These dendrograms were extracted in database format to proceed for RF and CART.

However, after formation of these natural data groups; identification of hazardous cluster was carried out at this stage. The result of HC is summarized in Table 5.1. Table 5.1 is self-explanatory. The red highlighted groups are identified as most hazardous clusters followed by yellow highlighted groups. The decisions are based on sample size and fatal percentage contribution in the groups. Due to low sample size multi vehicle accident clusters were discarded for CART (marked by red shades in Table 5.2) but included for RF. The final selection of clusters is summarized and highlighted with blue sheds in Table 5.2.

Table 5.1. Summary of hierarchical clustering

Accident Group	Cluster #	Accident Severity #						Sample %	Sample #	Accident Severity %				
		F	G	S	M	Total	Grand Total			F	G	S	M	Total
Pedestrian	Cluster_1	5749	987	138	NA	6874	7610	90	6874	84	14	2	NA	100
	Cluster_2	161	10	6	NA	177		2	177	91	6	3	NA	100
	Cluster_3	184	17	6	NA	207		3	207	89	8	3	NA	100
	Cluster_4	332	14	6	NA	352		5	352	94	4	2	NA	100
Double Vehicle	Cluster_1	1454	118	79	29	1680	5095	33	1680	87	7	5	2	100
	Cluster_2	808	508	74	80	1470		29	1470	55	35	5	5	100
	Cluster_3	832	477	103	367	1779		35	1779	47	27	6	21	100
	Cluster_4	126	27	8	5	166		3	166	76	16	5	3	100
Single Vehicle	Cluster_1	891	181	59	51	1182	1696	70	1182	75	15	5	4	100
	Cluster_2	90	52	11	22	175		10	175	51	30	6	13	100
	Cluster_3	114	34	22	0	170		10	170	67	20	13	0	100
	Cluster_4	147	15	4	3	169		10	169	87	9	2	2	100
Multi Vehicle	Cluster_1	1	10	4	3	18	61	30	18	6	56	22	17	100
	Cluster_2	17	1	2	0	20		33	20	85	5	10	0	100
	Cluster_3	14	1	0	4	19		31	19	74	5	0	21	100
	Cluster_4	0	3	1	0	4		7	4	0	75	25	0	100

Note: F= Fatal accident, G=Grievous accident, S=Simple injury accident, M=Motor collision/Property damage only (PDO) accident

Table 5.2. Cluster selection for CART

Accident Group	Cluster #	Accident Severity #						Sample %	Sample #	Non-fatal Accident #	Non-fatal Accident %	Accident Severity %				
		F	G	S	M	Total	Grand Total					F	G	S	M	Total
Pedestrian	Cluster_1	5749	987	138	NA	6874	7610	90	6874	1125	16	84	14	2	NA	100
	Cluster_2	161	10	6	NA	177		2	177	16	9	91	6	3	NA	100
	Cluster_3	184	17	6	NA	207		3	207	23	11	89	8	3	NA	100
	Cluster_4	332	14	6	NA	352		5	352	20	6	94	4	2	NA	100
Double Vehicle	Cluster_1	1454	118	79	29	1680	5095	33	1680	226	13	87	7	5	2	100
	Cluster_2	808	508	74	80	1470		29	1470	662	45	55	35	5	5	100
	Cluster_3	832	477	103	367	1779		35	1779	947	53	47	27	6	21	100
	Cluster_4	126	27	8	5	166		3	166	40	24	76	16	5	3	100
Single Vehicle	Cluster_1	891	181	59	51	1182	1696	70	1182	291	25	75	15	5	4	100
	Cluster_2	90	52	11	22	175		10	175	85	49	51	30	6	13	100
	Cluster_3	114	34	22	0	170		10	170	56	33	67	20	13	0	100
	Cluster_4	147	15	4	3	169		10	169	22	13	87	9	2	2	100
Multi Vehicle	Cluster_1	1	10	4	3	18	61	30	18	17	94	6	56	22	17	100
	Cluster_2	17	1	2	0	20		33	20	3	15	85	5	10	0	100
	Cluster_3	14	1	0	4	19		31	19	5	26	74	5	0	21	100
	Cluster_4	0	3	1	0	4		7	4	4	100	0	75	25	0	100

Note: F= Fatal accident, G=Grievous accident, S=Simple injury accident, M=Motor collision/Property damage only (PDO) accident

5.5 Application of RF

RF methodology was applied separately on all the hazardous clusters identified through HC. The randomForest package of R was used for this purpose. The intention was to identify the high impact variables individually from each of these datasets. The variables which topped the lists have been considered as most important predictors i.e. high impact variables (Appendix H). However, the predictors (variables) like number of vehicles involved, number of driver casualties, number of passenger casualties, and number of pedestrian casualties have been left out during summarizing the predictor (variable) selection for CART because of the obvious fact of their correlation with accident severity. Moreover, this dissertation is concerned with a national database and overall accident severity pattern i.e. how accident severities are related to roads and roadway environment and operating condition is most important in this study. Therefore, the predictors related to these issues have been inspected with greater emphasis than temporal variables like time, date, month, and year of accident events in CART; although during RF these temporal variables have been considered with the same gravity like all other variables.

The predictors have been selected based on their variable importance as outlined in Chapter 3. Variable importance is a difference and this difference is measured considering modulus sign. In the R program variable importance is called as mean decrease accuracy, and is measured without considering the modulus sign. Therefore, during predictor selection, this issue was taken into consideration. The selected final variables through RF for CART are summarized in Table 5.3 along with their mean decrease accuracy. For easy understanding the predictors are highlighted with green shades. Moreover, the description of theses predictors is abridged in Table 5.4.

Table 5.3. Summary of RF

Accident Group	Cluster #	High Impact Variables				
Pedestrian	Cluster_1	RdClassM	LightM	CollTypeM	LocatTypeM	DividerM
	Mean Decrease Accuracy	26.9062199	26.83743161	23.72993413	22.5120291	16.58498867
	Cluster_2	LocatTypeM	TrafficContrlM	RdClassM	JuncTypeM	NA
	Mean Decrease Accuracy	8.934941283	5.746201222	4.269378045	4.148095324	NA
	Cluster_3	TrafficContrlM	Rd_GeoM	CollTypeM	JuncTypeM	NA
	Mean Decrease Accuracy	4.783570879	-2.06444957	-1.739744352	1.733000617	NA
	Cluster_4	JuncTypeM	Surf_TypeM	RdClassM	TrafficContrlM	NA
	Mean Decrease Accuracy	9.237595977	-4.32681388	3.237877133	3.188050255	NA
Double Vehicle	Cluster_1	Rd_GeoM	LightM	CollTypeM	Surf_TypeM	TrafficContrlM
	Mean Decrease Accuracy	10.30768083	6.152270616	6.039192418	5.525508339	4.80883211
	Cluster_4	RdClassM	CollTypeM	Surf_CondM	LocatTypeM	Rd_GeoM
	Mean Decrease Accuracy	3.184766888	3.148147297	2.011488866	1.615001565	1.532404265
Single Vehicle	Cluster_1	CollTypeM	LightM	Rd_GeoM	TrafficContrlM	JuncTypeM
	Mean Decrease Accuracy	8.697937388	6.407632429	6.195833895	4.119314096	3.779245973
	Cluster_3	Surf_TypeM	LightM	JuncTypeM	WeatherM	NA
	Mean Decrease Accuracy	4.09891818	3.497018816	3.413011471	2.282952542	NA
	Cluster_4	CollTypeM	MovM	Surf_CondM	NA	NA
	Mean Decrease Accuracy	5.53668368	4.982673333	3.452905007	NA	NA
Multi Vehicle	Cluster_2	WeatherM	LocatTypeM	Surf_CondM	Rd_GeoM	NA
	Mean Decrease Accuracy	-1.9806105	1.051854514	0.926921057	0.229327544	NA
	Cluster_3	JuncTypeM	LightM	LocatTypeM	Rd_GeoM	NA
	Mean Decrease Accuracy	-3.304094153	2.992148949	-2.172437226	1.784178541	NA

Table 5.4. Description of variables

Variable (Predictor)	Variable Description
No_VehInv	Number of Vehicles Involved
No_DrvCasu	Number of Driver Casualties
No_PassCasu	Number of Passenger Casualties
No_PedCasu	Number of Pedestrian Casualties
Day	Day of Week
Month	Month
Time_SQL	Time of Accident Occurrence
JuncTypeM	Junction Type
TrafficContrlM	Traffic Control Type
CollTypeM	Collision Type
MovM	One-way or Two-way Movement
DividerM	Presence of Divider
WeatherM	Weather Condition
LightM	Light Condition
Rd_GeoM	Road Geometry
Surf_CondM	Surface Condition
Surf_TypeM	Surface Type
Surf_QualM	Surface Quality
RdClassM	Road Class
RdFeatuM	Road Feature
LocatTypeM	Location Type

5.6 Application of CART

Classification and regression trees (CART) partition the entire data space into small manageable chunks which facilitates clear understanding of problem domains. Therefore, classification trees have been grown using CART methodology with the datasets that have been identified as most critical clusters through hierarchical clustering (HC). Important predictors as identified by random forest (RF) have been given the priority during applying CART methodology. However, pedestrian accidents, double vehicle accidents and single vehicle accidents have been treated separately as before. Multi vehicle accident records in separate clusters are quite insignificant (because of low sample size) to apply CART on these clusters. So these clusters have not been selected for this analysis. The predictors like number of vehicles involved, number of driver casualties, number of passenger casualties, and number of pedestrian casualties have been left out during predictor selections for CART because of the obvious fact of their correlation with accident severity as outlined in RF; although they have been given the same gravity like other predictors during the application of RF.

This study is concerned about how accident severities are related to road and roadway environment and operating conditions. Therefore, the predictors related to these issues have been inspected with greater emphasis than temporal variables like time, date, month and year of the accident events; besides, temporal variables are more useful to get deep insight into crashes in specific geographical locations and geocoded data would become handy in analyzing the status. Moreover, instead of growing a tree up to maximum depth where each terminal node comprises a pure class, a minimum split rule has been used in such a way that a node gets split only when it contains at least 10 data points so that subsequent child nodes have at least $1/3^{rd}$ of those data points. This facilitates in reducing the calculation complexity as well as the tree size substantially. Furthermore, from analytical point of view it infers insight into the situation easily. If the parent node is ‘n’ then the left and the right child nodes are numbered as ‘2n’ and '2n+1' respectively.

5.6.1 Pedestrian accidents

The whole dataset of pedestrian accidents have been divided into four major clusters through HC. Among these, Cluster 1 is the most hazardous cluster as outlined earlier. Therefore, Cluster 1 of pedestrian accidents has been analyzed first in the following. Then the other clusters of pedestrian accidents were analyzed by CART.

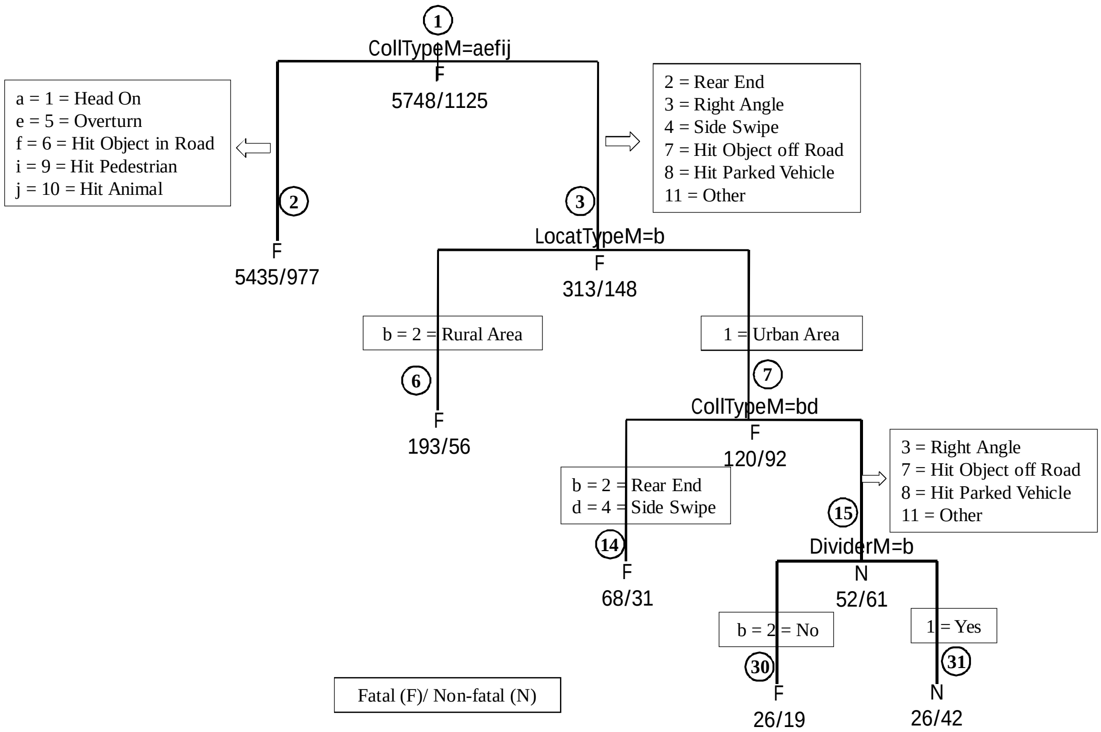


Figure 5.1. Classification tree for pedestrian accidents cluster 1.

Collision type emerges as the splitting predictor of the root node. It is quite interesting to find that pedestrians are not only the direct victims of hit pedestrian accidents; rather all other accident types are contributing to the pedestrian vulnerability in Bangladesh. This also highlights the fact that probably pedestrians are not getting the proper pedestrian facilities and are widely exposed throughout the road networks. However, hit pedestrian along with head on, overturn, hit object in road and hit animal collision types contribute 84.76 percent fatal pedestrian accidents as outlined in Node 2 of Figure 5.1. Rear end, right angle, side swipe, hit object off road, hit parked vehicle and other collision types contribute less fatal (67.89%) pedestrian accidents (Node 3). Moreover, in the second collision (Node 3) group right angle, hit object off road, hit parked vehicle and other collision types are more involved in non-fatal accidents (52 fatal versus 61 non-fatal, Node 15). Dividers in urban areas have a positive impact in reducing fatal pedestrian accidents. Urban divided roadways have less fatal pedestrian accidents (38.23%, node 31) compared to fatal pedestrian accidents (57.77%, node 30) on urban undivided roadways. On the other hand, rural areas are substantially hazardous for pedestrians as depicted in Node 6 (193 fatal pedestrian accidents versus 56 non-fatal pedestrian accidents).

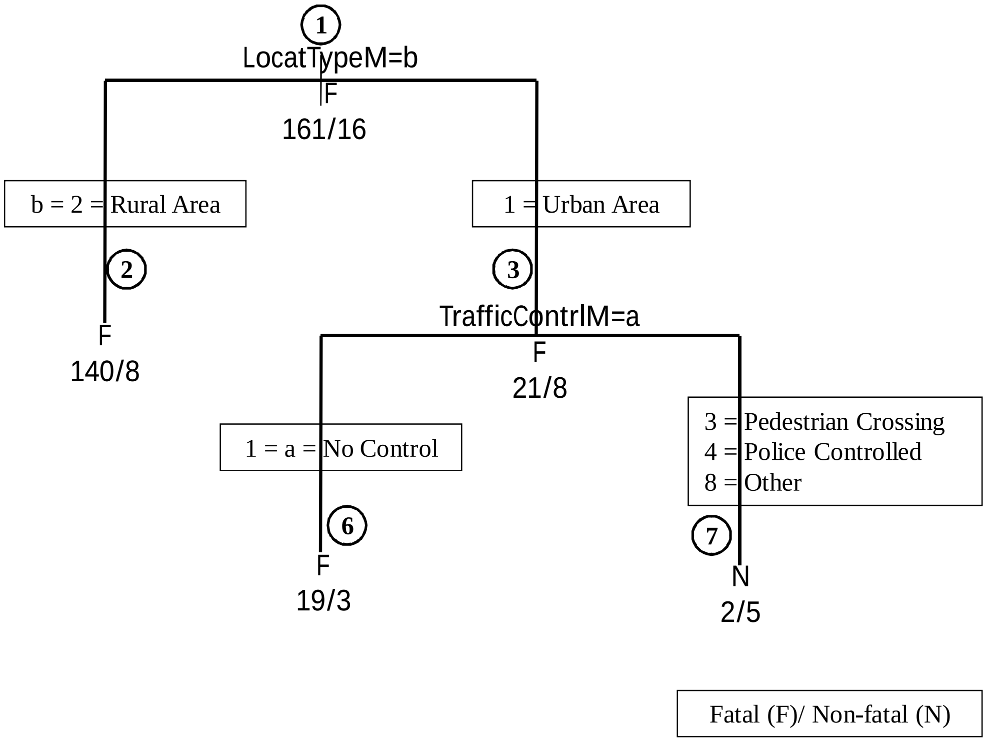


Figure 5.2. Classification tree for pedestrian accidents cluster 2.

CART provides another stimulating fact about pedestrian accidents in Cluster 2 (Figure 5.2). More than 85 percent fatal pedestrian accidents are clustered in the roadway sections where no traffic control system is available. On the contrary, only 28.57 percent fatal pedestrian accidents occur where there is some form of traffic control system, viz., pedestrian crossing, police control, etc. (Node 7, Figure 5.2) are available. And it is to be noted that this is the scenario of urban areas. This clearly implies the need for proper traffic control measures. Figure 5.2 also exhibits pedestrian to be more susceptible to fatal accidents (Node 2, 94.59% fatal) on the rural sections of our road network as like Cluster 1. This demonstrates the possibility of rural roads having more exposure which leads to more crashes resulting in higher number of fatal accidents.

Traffic control predictor splits the root node of Cluster 3 for pedestrian accident dataset. Analogous to cluster two, cluster three illustrated that in case of police controlled traffic system no fatal pedestrian accident has been recorded (Node 3, Figure 5.3). Other traffic control systems viz. no control, centerline marking, pedestrian crossing have been underlined with 89.75 percent fatal pedestrian accidents; and the major percentage of this representation is due to hit pedestrian along with head on, rear end, overturn and other types of collision (Node 4, Figure 5.3). Fatal pedestrian accidents are more prone to straight and flat and curve road geometric sections (Node 10, 100% fatal pedestrian accidents) than slope only geometric sections (Node 11) and this consequence has been associated with side swipe and hit object off road collision types and where there is no police controlled traffic system available. Therefore, it can be easily perceived that traffic control system and road geometry have definite impact on pedestrian accident severity.

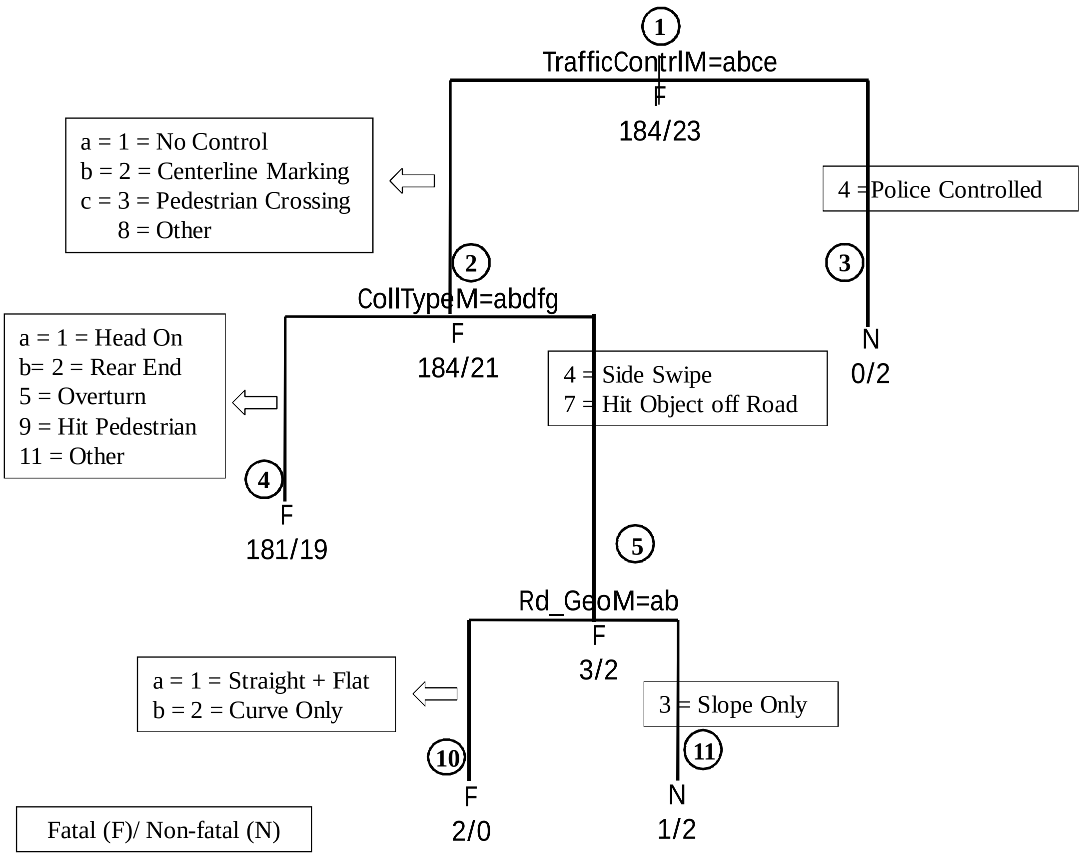


Figure 5.3. Classification tree for pedestrian accidents cluster 3.

It has been identified from Cluster 4 of pedestrian accidents that national highways' cross junctions with no or centerline marking traffic control system are highly vulnerable sections for pedestrians (Node 14, Figure 5.4). Rural and city roads cross junctions are safer (46.15% non-fatal pedestrian accidents) platform for pedestrians with no or centerline marking traffic control system compared to that of national highways. But centerline marking traffic control system hardly had any impact on reducing pedestrian accident severity on rural and city roads (Node 30 and Node 31 of Figure 5.4). Unlike Cluster 3, at cross junctions police controlled traffic system has been found ineffective in reducing fatal pedestrian accidents (Node 6 of Figure 5.4, 87.5% fatal pedestrian accidents). The mid-section of roadways and junction types like tee, staggered, roundabout and railway crossings have been underscored with 96.47 percent fatal pedestrian accidents (Node 2); which depicted the fact that pedestrian accidents are spread throughout the road network and lack of proper pedestrian facilities could be one the main factors behind this carnage.

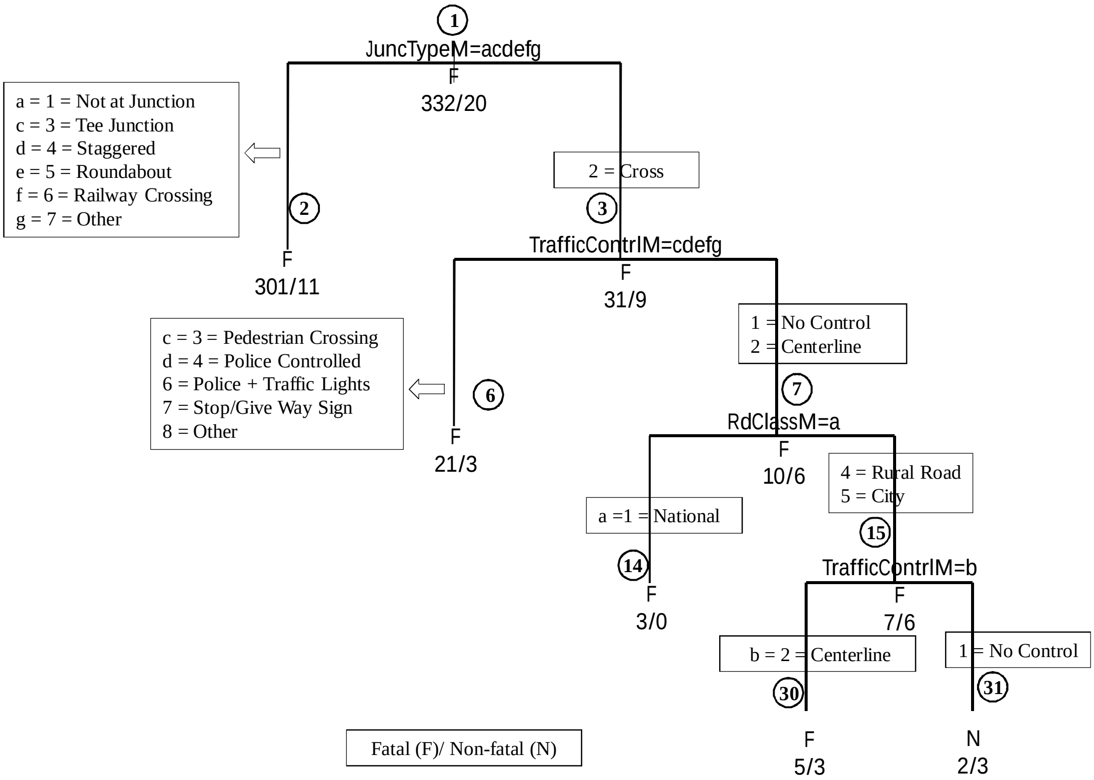


Figure 5.4. Classification tree for pedestrian accidents cluster 4.

5.6.2 Double vehicle accidents

Road geometry has differentiated the root node of Cluster 1 for double vehicle accidents (Figure 5.5). It has been found that more fatal accidents (89.03%) have taken place in straight and flat, slope only, and crest oriented geometric sections of roads (Node 2, Figure 5.5) compared to curve only, and curve and slope oriented geometric sections (Node 3 of Figure 5.5, 72.17% fatal accidents). However, lighting condition associated with collision type has a significant impact on the latter part of the previous finding. It has been identified that if the lighting condition is dawn or dusk and night-unlit (Node 12, Figure 5.5) and if the accident collision type is head on, right angle, overturn, hit object in road, and hit animal (Node 6, Figure 5.5) then 90.90 percent double vehicle accidents resulted in fatal cases (Node 12, Figure 5.5).

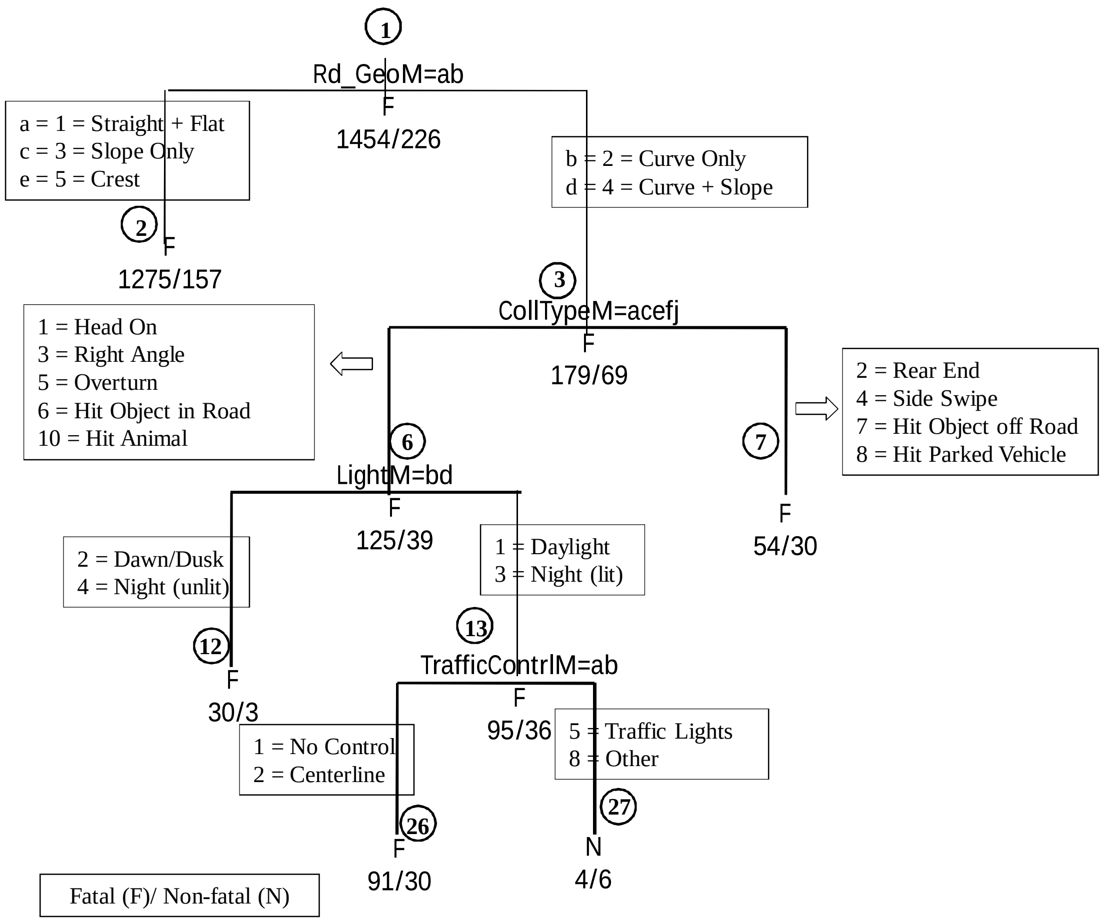


Figure 5.5. Classification tree for double vehicle accidents cluster 1.

But with the same collision types if the accidents occurred in daylight and night-lit condition concomitant with traffic lights traffic control system then 60 percent of these accidents stemmed in non-fatal crashes (Node 27, Figure 5.5); which is quite the reverse case (75.20% fatal accident) if there is no traffic control system and centerline marking traffic control system (Node 26, Figure 5.5). Moreover, rear end, side swipe, hit object off road, and hit parked vehicle collision type double vehicle accidents occurred on curve only, and curve and slope sections of roads ensued in 64.28 percent fatal crashes.

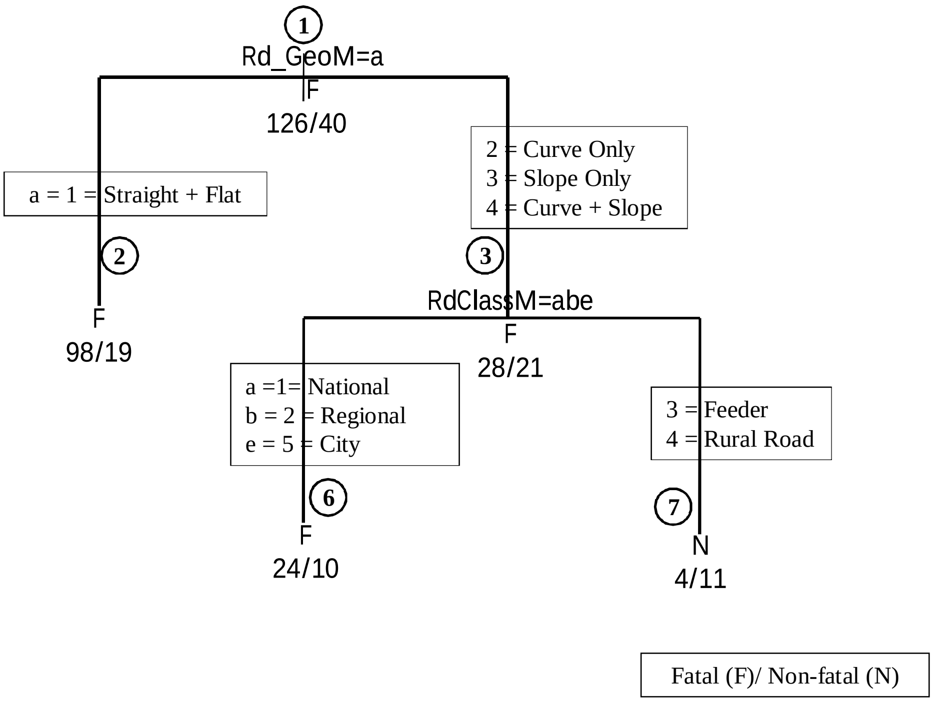


Figure 5.6. Classification tree for double vehicle accidents cluster 4.

Cluster 4 for double vehicle accidents has braced the fact that fatal accidents are more susceptible to straight and flat portion of roadways (83.76% fatal accidents, Node 2 of Figure 5.6). On the contrary, accidents on feeder and rural roads allied with curve only, slope only, and curve and slope geometric segments ended up in 73.33 percent non-fatal cases (Node 7, Figure 5.6); but for national, regional, and city road segments with the same affiliation the opposite is found true (70.59% fatal accidents, Node 6 of Figure 5.6).

5.6.3 Single vehicle accidents

Collision type transpires as the splitting predictor of the root node for single vehicle Cluster 1 dataset. This first split has highlighted a feeble point of the accident database as well. It is found that even in single vehicle accident database head on and rear end accidents have been registered. Nevertheless, rear end, overturn, hit parked vehicle, hit animal, and other types of accidents constituted 80 percent single vehicle fatal accidents (Node 2, Figure 5.7). On the other hand, head on, right angle, side swipe, hit object in road, and hit object off road comprised 167 fatal accidents versus 110 non-fatal accidents (Node 3, Figure 5.7). Curve only, slope only, and curve and slope oriented geometric sections associated with the latter part of the above mentioned collision group have been found clustered with 85.29 percent fatal accidents (Node 6, Figure 5.7). However, when the sections had been straight and flat, and crest 56.79 percent fatal crashes have been registered (Node 7, Figure 5.7).

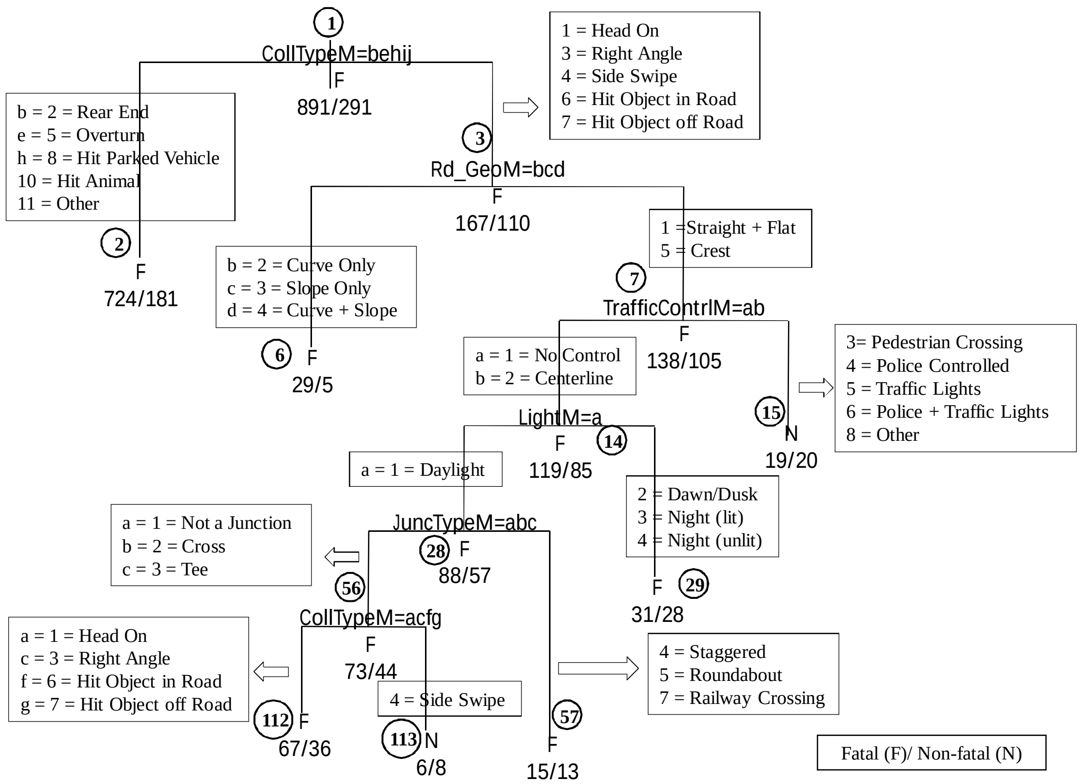


Figure 5.7. Classification tree for single vehicle accidents cluster 1.

Node 15 of Figure 5.7 revealed that pedestrian crossing, police control, traffic lights, police control and traffic lights have some positive impacts in reducing the fatal accidents a little bit (19 fatal versus 20 non-fatal crashes); and this is true for head on, right angle, side swipe, hit object in road, and hit object off road collision types along with straight and flat, and crest road segments. Nonetheless, no traffic control, and centerline marking traffic control system have been found quite ineffective (58.33% fatal crashes, Node 14 of Figure 5.7) in the above mentioned cases. Node 28 has depicted that daylight manifested more fatal accidents (Figure 5.7, 88 fatal versus 57 non-fatal) than dawn or dusk, night-lit, and night-unlit lighting conditions (Node 29, 31 fatal versus 28 non-fatal); and this is associated with all other conditions of Node 14 of Figure 5.7. Accidents occurring at mid-block sections, cross junctions, and tee junctions are more prone to fatal cases (Node 56 of Figure 5.7, 62.39% fatal) than accidents occurring at staggered junctions, roundabouts, and railway crossings (Node 57 of Figure 5.7, 53.57% fatal).

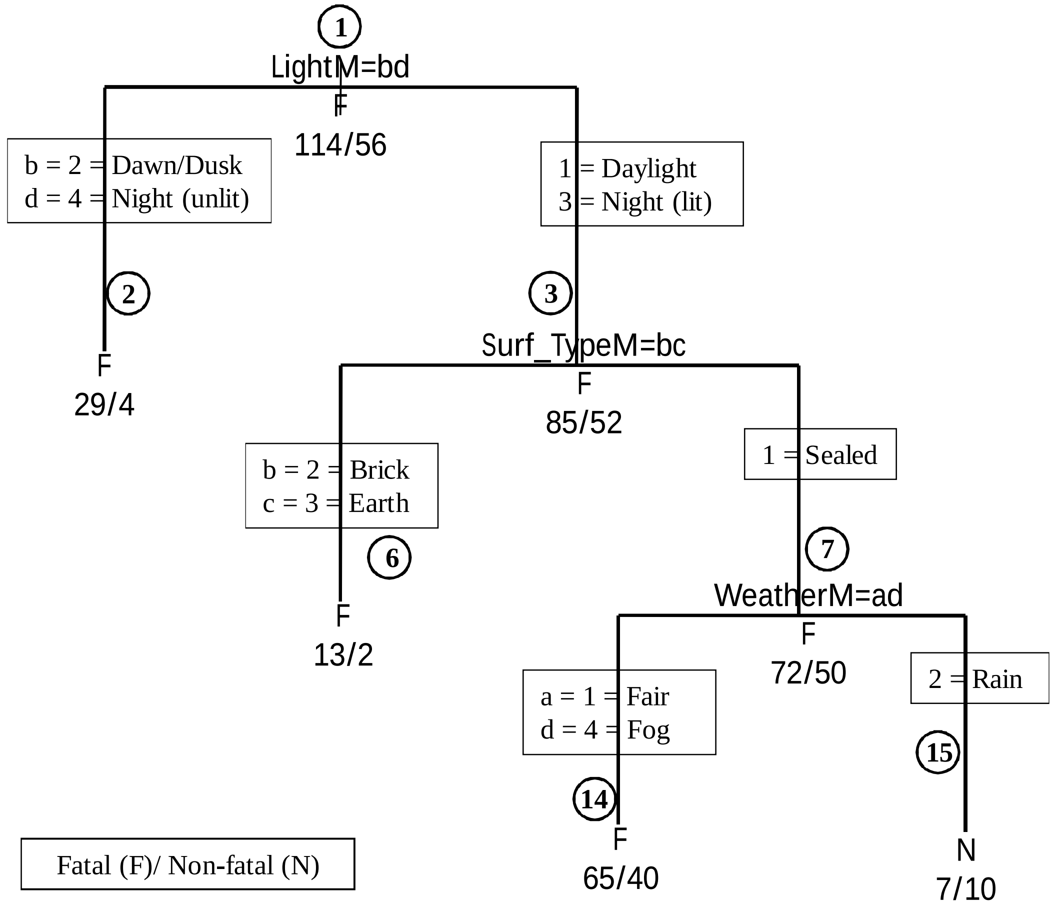


Figure 5.8. Classification tree for single vehicle accidents cluster 3.

Cluster 3 for single vehicle accidents outlined a different view than Cluster 1 regarding lighting condition. Dawn or dusk, and night-unlit lighting condition have been identified concomitant with more fatal crashes (29 fatal versus 4 non-fatal, Node 2 of Figure 5.8) than daylight, and night-lit lighting condition (62.04% fatal, Node 3 of Figure 5.8). Furthermore, brick and earthen surface type along with daylight and night-lit lighting condition have been underscored with 13 fatal crashes versus 2 non-fatal crashes (Node 6, Figure 5.8). Interestingly rainy weather allied with sealed surface and daylight and night-lit lighting situation came up with less fatal crashes (10 non-fatal versus 7 fatal, Node 15 of Figure 5.8) than the same conditions with fair and foggy weather (61.90% fatal accidents, Node 15 of Figure 5.8).

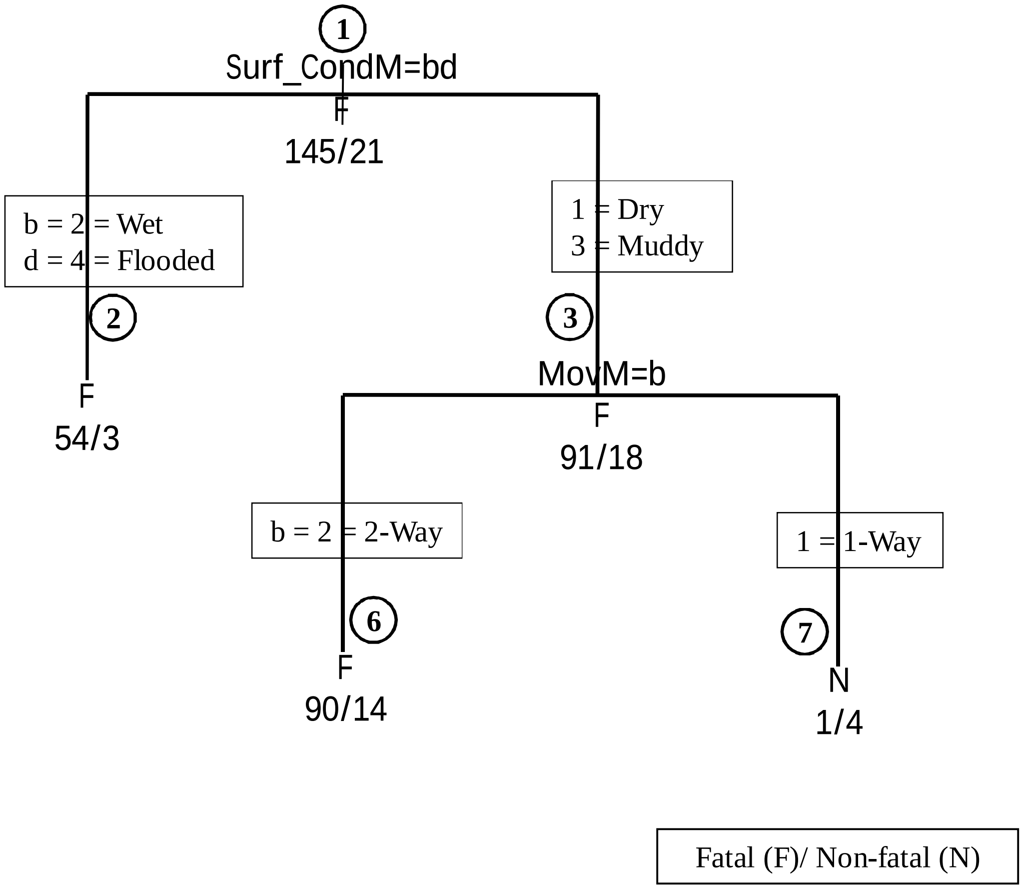


Figure 5.9. Classification tree for single vehicle accidents cluster 3.

CART provides another stimulating fact about single vehicle accidents through Cluster 4. The root node of this cluster has been differentiated with road surface condition variable. It appears that wet and flooded surface condition contributed in huge fatal accidents (94.73%, Node 2 of Figure 5.9). Whereas dry and muddy surface along with two-way roads suffered with 86.53 percent fatal accidents (Node 6, Figure 5.9). And as was expected one-way roads even with dry and muddy surface conditions have contributed to only 20 percent of fatal crashes for single vehicle accidents (Node 7, Figure 5.9).

CHAPTER 6
CONCLUSION

6.1 General

This thesis is the first step regarding the application of data mining (DM) in road traffic accident analysis in Bangladesh. No previous studies have ever utilized data mining for finding unrecognized and unsuspected facts and overcome road accident data related deficiencies as well as statistical limitations in the country. Therefore, in this research, an attempt has been made to study the feasibility and utility of data mining methods in the context of road traffic safety of Bangladesh. The intention was to elicit reasonable, and hopefully novel, unsuspected and interesting facts as well as confirming any perceived concepts from road traffic accident data. This chapter mainly summarizes the findings of this research and outlines the precincts and future research scopes.

6.2 Key Findings

The primary finding of this study is that DM has depicted few practical, unique, unanticipated, and attention-grabbing realities as well as has confirmed some perceived facts from road traffic accident database. It has been able to quantify multiple predictors' relationships which lead to crashes and eventual injuries in accident events. These facts and relationships can lead to better understanding of the accident phenomena where traditional statistical approaches have failed to instigate so far; i.e. DM is capable of dealing with large datasets and drawing pragmatic conclusions where as traditional approaches involve human exploration tasks, and thereby limits the assessment capacity. DM has been able to overcome the limitations of traditional approaches regarding road traffic accident analysis, and thus designing proper countermeasures and policy level decisions. Following are the key findings of this thesis:

- Nearly 10 percent of the pedestrian accidents are triggered by other accident/collision types, which indicate that may be pedestrians are not only the victims but also a stimulating factor for some accidents.
- Pedestrians are found to be more vulnerable in rural areas (77.5% fatal accidents) than in urban areas (56.6% fatal accidents) as secondary dupes. Exposure might be a reason but it also highlights the fact that probably pedestrians are not getting proper pedestrian facilities and are widely exposed throughout the road network.
- Dividers in urban areas have been found quite effective in reducing fatal (38.23% fatal vs 57.78% fatal where there are no dividers) pedestrian accidents.
- Traffic control systems especially police controlled traffic control system in urban areas have been identified as persuasive in reducing pedestrian fatal accidents (in some cases 0% fatal incidences).
- Geometric sections without police controlled traffic control system have been acknowledged as a bracing factor for fatal pedestrian accidents. The straight and flat, and curve only geometric sections associated with side swipe and hit object off road type accidents provoked more fatal pedestrian accidents (nearly 100% fatal compared to 33.33% fatal in case of slope only geometric sections).
- National highways' cross junctions without any traffic control system or with centerline marking traffic control system are highly vulnerable sections for pedestrians. However, rural and city roads cross junctions with the same aspects are safer (100% vs 53.85% fatal pedestrian accidents).
- The mid-section of roadways and junction types like tee, staggered, roundabout, and level crossings have been underscored with more fatal pedestrian accidents than cross junctions.
- Straight and flat geometric sections of roadways have generated more double vehicle fatal accidents (more than 80% accidents are fatal) than other types (e.g. curve only, slope only, curve and slope and crest) of geometric sections (nearly 70% fatal accidents).
- The latter part of the previous finding got worse when the sections were associated with head on, right angle, overturn, hit object in road and hit animal type of collisions (76.22% fatal); or occurred on national and regional highways or feeder roads (71% fatal); or during dawn/dusk and night (unlit) lighting

condition (90.91% fatal); or in daylight or night (lit) light condition but with no or centerline marking traffic control system (75.21% fatal).

- It has been found that head on and rear end collisions are recorded in the single vehicle accident database which transpires a feeble point of the accident database.
- However, rear end, overturn, hit parked vehicle and hit animal type accidents constituted 80 percent fatal accidents for single vehicle crashes.
- On the other hand, head on, right angle, side swipe, hit object in road, and hit object off road collision types affiliated with curve only, slope only, and curve and slope geometric sections of the roadways produced 85.29 percent fatal single vehicle crashes.
- Even straight and flat, and crest geometric sections allied with no and centerline marking traffic control system induced 58.33 percent single vehicle fatal accidents.
- Dawn/dusk and night (unlit) lighting condition attributed 87.88 percent single vehicle fatal accidents. Even daylight akin with head on, right angle, hit object in road and hit object off road type collisions at mid-block sections of roads, and at cross and tee junctions has resulted in 65.05 percent fatal single vehicle accidents.
- Staggered junctions, roundabouts and level crossings have been identified responsible for 46.43 percent non-fatal single vehicle crashes. Pedestrian crossings, police control, and traffic lights have been underscored with some persuasive impacts on reducing fatal single vehicle accidents (51.28% non-fatal accidents) even in straight and flat geometric sections of roads.
- Brick and earthen road surfaces have generated 86.67% fatal single vehicle crashes even in daylight and night (lit) condition. On the contrary, sealed surface even affiliated with rainy weather has ensued less fatal single vehicle crashes (58.82% non-fatal crashes).
- Wet and flooded surface conditions of roads have resulted in 94.74 percent fatal single vehicle crashes. Nevertheless, one-way routes concomitant with dry and muddy surface prompted only 20 percent fatal cases as always perceived; whereas in case of two-way roads it shoots up to 86.54 percent fatal single vehicle accidents.

The findings clearly demonstrate the capability of data mining in making pragmatic transportation decisions and allow us to allocate resource accordingly. This cannot be achieved by general statistical tools. Statistics may help us to quantify a particular issue but cannot give us the insight, quantify multiple relationships, and make policy level decisions i.e. where to apportion the budget to reduce what percentage of loss. Data mining provides platforms (hypotheses) that are beyond human exploration task which is the utmost need in making financial decisions. Therefore, the feasibility and utility of data mining are justified in the context of Bangladesh's road safety status.

6.3 Limitations of the Study

Resource limitations had been one of the most important determinants for this research. Processing the massive accident database for data mining requires fast computers with sound technical configuration. It was found that the computer that was designated for the data mining operations for this dissertation could not handle the huge database. Therefore, it became urgent to reduce the size of the accident database. It is expected that if it would have been possible to work with the whole database the results might have been more precise.

A tool was required to convert/transcript the whole MAAP database to MS Excel smoothly and correctly. Then it would have been possible to encompass the other predictors in this study that have been left out due to inconsistency in the column heads.

The number of clusters produced in HC and depth of classification trees grown in CART were controlled to reduce the complexity of the study and for easy understanding of the problem domains. As this dissertation has been the first step towards the application of data mining in road traffic accident analysis in Bangladesh, simplicity has been endured in back of mind.

6.4 Future Research Scopes

It is being perceived that by overcoming the study limitations new research horizons would be yielded. These can be abridged as follows:

- The DM process was executed through a laptop configured with Intel core i5 processor and 6GB RAM. Future researches should lodge better PCs to incorporate the complete road traffic accident database.
- A conversion toll should be developed to transcript the MAAP database to easy importable database for R so that all the predictors could be assimilated in the studies.
- Number of clusters can be increased through HC and depth of CART trees can be enlarged in future works to get more micro-level aspects clearly. However, this would be a time consuming process as well.
- Multi vehicle accident database was discarded in CART phase due to low sample size in each of the 2 clusters. Two things can be done in future from this database; either it can be merged with double vehicle accident database or direct CART can be executed without processing the database with HC.
- This study has been concerned with how accident severities are related to road and roadway environment and operating conditions, i.e. the target/dependent predictor was accident severity. However, it is possible to change the target predictor to any other variables like accident/collision type, road class, etc. and draw new relationships accordingly.
- Weight or gravity of the variables/predictors during the application of hierarchical clustering (HC) was considered same. Nevertheless, different weights can be assigned to different predictors to stimulate policy level decisions.
- In-depth analyses of data mining findings for developing countermeasures and policy level decisions would provide enormous scopes for future endeavors.

REFERENCES

Abdalla, I. M., and Robert, R., "An investigation into the relationships between area social characteristics and road accident casualties," *Accidents Analysis and Prevention*, Vol. 5, pp. 583–593, 1997.

Abdel-Aty, M., Dilmore, J., and Dhindsa, A., “Evaluation of variable speed limits for real-time freeway safety improvement,” *Accident Analysis and Prevention*, Vol. 38(2), pp. 335–345, 2006a.

Abdel-Aty, M., Pemmanaboina, R., and Hsia, L., “Assessing crash occurrence on urban freeways by applying a system of interrelated equations,” *Transportation Research Record*, Vol. 1953, pp. 1–9, 2006b.

Abdel-Aty, M., Uddin, N., and Pande, A., “Split models for predicting multivehicle crashes during high-speed and low-speed operating conditions on freeways,” *Transportation Research Record*, Vol. 1908, pp. 51–58, 2005.

Abdel-Aty, M., and Abdalla, F., “Linking roadway geometrics and real-time traffic characteristics to model daytime freeway crashes using generalized extreme equations for correlated data,” *Transportation Research Record*, Vol. 1897, pp. 106–115, 2004.

Abdel-Aty, M., and Pande, A., “Classification of real-time traffic speed patterns to predict crashes on freeways,” in *2004 Proceedings of the 83rd Annual Meeting of Transportation Research Board*, Washington, DC, 2004.

Alam, M. J. B., Karim, D. M., Hoque, M. M., Islam, Q. Z., and Alam, M. D., “Initiatives regarding road accident database in Bangladesh,” in *2006 International Conference on Road Safety in Developing Countries*, Aug. 2006, pp. 253-258.

Barai, S. K., “Data Mining Applications in Transportation Engineering,” Dept. of Civil Engineering, Indian Institute of Technology Kharagpur, 2003.

http://www.tandfonline.com/doi/abs/10.1080/16483840.2003.10414100#preview accessed on May 29, 2013.

Bayam, E., Liebowitz, J., and Agresti, W., “Older drivers and accidents: A meta analysis and data mining application on traffic accident data,” in *2005 Proceedings of 29 Expert Systems with Applications*, 2005, pp. 598–629.

Bedard, M., Guyatt, G. H., Stones, M. J., and Hireds, J. P., “The Independent Contribution of Driver, Crash, and Vehicle Characteristics to Driver Fatalities,” *Accident analysis and Prevention*, Vol. 34, pp. 717–727, 2002.

Beshah, T., “Application of data mining technology to support RTA severity analysis at Addis Ababa traffic office,” Addis Ababa University, Addis Ababa, 2005.

Beshah, T., and Abraham, A., "Rule Mining and Classification of Road Traffic Accidents Using Adaptive Regression Trees," *Journal of Simulation*, Vol. 6, pp. 10–11, 2005.

Breiman, L., “Random Forests,” *Journal of Machine Learning*, Vol. 45(1), pp. 5-32, 2001.

Breiman, L., “Bagging Predictors,” *Journal of Machine Learning*, Vol. 24(2), pp. 123-140, 1996.

Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984) *Classification and Regression Trees*. Wadsworth Inc., Pacific Grove, CA.

Caliendo, C., Guida, M., and Parisi, A., “A crash-prediction model for multilane roads,” *Accident Analysis and Prevention*, Vol. 39(4), pp. 657–670, 2007.

Cameron, M., “Accident data analysis to develop target groups for countermeasures,” Monash University Accident Research Centre, Australia, Reports 46 & 47, 1997.

Cedar, A., and Livneh, M., “Relationship between road accidents and hourly traffic flow – I,” *Accident Analysis and Prevention*, Vol. 14(1), pp. 19–34, 1982.

Cedar, A., “Relationship between road accidents and hourly traffic flow – II,” *Accident Analysis and Prevention*, Vol. 14(1), pp. 35–44, 1982.

Chang, L., and Wang, H., "Analysis of traffic injury severity: An application of non-parametric classification tree techniques,” *Accident Analysis and Prevention,* Vol. 38(5), pp. 1019–1027, 2006.

Chang, L., and Chen, W., “Data mining of tree-based models to analyze freeway accident frequency,” *Journal of Safety Research*, Vol. 36, pp. 365–375, 2005.

Chen, H., Zhou, H., Zhao, J., and Hsu, P., “Safety performance evaluation of left-side off-ramps at freeway diverge areas,” *Accident Analysis and Prevention,* Vol. 43(3), pp. 605–612, 2010.

Chen, H., Liu, P., Lu, J.J., and Behzadi, B., “Evaluating the safety impacts of the number and arrangement of lanes on freeway exit ramps,” *Accident Analysis and Prevention,* Vol. 41(3), pp. 543–551, 2009.

Chen, W., and Jovanis P., “Method for identifying factors contributing to driver-injury severity in traffic crashes,” *Transportation Research Record,* Vol. 1717, pp. 1–9, 2002.

Chong, M., Abraham, A., and Paprzycki, M., “Traffic accident analysis using machine learning paradigms,” *Informatica,* Vol. 29, pp. 89–98, 2005.

Christoforou, Z., Cohen, S., and Karlaftis, M.G., “Integrating real-time traffic data in road safety analysis,” in *2012 Proceedings of Social and Behavioral Science,* Vol. 48, 2012, pp. 2454–2463.

Christoforou, Z., Cohen, S., and Karlaftis, M.G., "Identifying crash type propensity using real-time traffic data on freeways," *Journal of Safety Research*, Vol. 42, pp. 43–50, 2011.

Clarke, R., Forsyth, R., and Wright, R., "Machine learning in road accident research: decision trees describing road-accidents during cross-flow turns," *Ergonomics*, Vol. 41(7), pp. 1060–1079, 1998.

Dalgaard, P. (2008) *Introductory Statistics with R*. Springer, NY.

Dias, C., Miska, M., Kuwahara, M., and Warita, H., "Relationship between congestion and traffic accidents on expressways: an investigation with Bayesian belief networks," in *2009 Proceedings of 40th Annual Meeting of Infrastructure Planning (JSCE)*, Japan, 2009.

Dia, H., and Rose, G., "Development and Evaluation of Neural Network Freeway Incident Detection Models Using Field Data," *Transportation Research*, Vol. 5(5), pp. 313–331, 1997.

Ekhaus, M. A., "Conquering the Differences between Data Mining and Statistics," Part I, Gibraltar Analytics, 2003. http:// www.gibraltaranalytics.com/docs/stats-dm.pdf accessed on March 8, 2012.

Estivill-Castro, V., and Yang, J., "A Fast and robust general purpose clustering algorithm," in *2000 Pacific Rim International Conference on Artificial Intelligence*, 2000, pp. 208-218.

Evanco, W. M., "The Potential Impact of Rural Mayday Systems on Vehicular Crash Fatalities," *Accident Analysis and Prevention*, Vol. 31, pp. 455–462, 1999.

Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., and Uthursamy, R. (1996) *Advances in Knowledge Discovery and Data Mining*. AAAI Press/The MIT Press, Cambridge, MA.

Fraley, C., and Raftery, A. E., "How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis," Department of Statistics, University of Washington, Tech. Rep. 329, 1998.

Frantzeskakis, J. M., and Iordanis, D. I., "Volume-to-capacity ratio and traffic accidents on interurban four-lane highways in Greece," *Transportation Research Record*, Vol. 1112, pp. 29–38, 1987.

Fridstrom, L., Ifver, J., Ingebrigtsen, S., Kulmala, S.R., and Thomsen, L.K., "Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts," *Accident Analysis and Prevention*, Vol. 27(1), pp. 1–20, 1995.

Friedman, J. H., "Data Mining and Statistics: What's the Connection?," Department of Statistics and Stanford Linear Accelerator Center, Stanford University, Stanford, CA 94305, 1997. http://www-stat.stanford.edu/~jhf/ftp/dm-stat.pdf accessed on May 29, 2013.

Geurts, K., Thomas, I., and Wets, G., "Understanding spatial concentrations of road accidents using frequent item sets," *Accident Analysis and Prevention*, Vol. 37(4), pp. 787–799, 2005.

Geurts, K., Wets, G., Brijs, T., and Vanhoof, K., "Profiling high frequency accident locations using association rules," in *2003 proceedings of the 82th Annual Meeting of the Transportation Research Board*, Washington, D.C., 2003.

Getnet, M., "Applying data mining with decision tree and rule induction techniques to identify determinant factors of drivers and vehicles in support of reducing and controlling road traffic accidents: the case of Addis Ababa city," Addis Ababa University, Addis Ababa, 2009.

Golob, T.F., Recker, W.W., and Alvarez, V.M., "A tool to evaluate the safety effects of changes in freeway traffic flow," in *2003 Proceedings of the 82nd Annual Meeting* *of Transportation Research Board*, Washington, DC, 2003.

Guha, S., Rastogi, R., and Shim, K., “CURE: An efficient clustering algorithm for large databases,” in *1998 ACM SIGMOD International Conference on Management of Data*, New York, 1998, pp. 73-84.

Han, J., and Kamber, M. (2001) *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, San Francisco, CA.

Hand, D., Mannila, H., and Smyth, P. (2001) *Principles of Data Mining*. The MIT Press, Massachusetts.

Harb, R., Yan, X., Radwan, E., and Su, X., “Exploring Precrash Maneuvers Using Classification Trees and Random Forests,” *Journal of Accident Analysis and Prevention*, Vol. 41(1), pp. 98-107, 2009.

Hill, T., and Lewicki, P. (2006) *Statistics: Methods and Applications, A Comprehensive Reference for Science, Industry, and Data Mining*. StatSoft Inc., Tulsa, OK.

Hoque, M. M., Smith, G., Hossain, D. Z., and Mahmud, S. M. S., “Improving Highway Safety in Bangladesh: Road Improvement and The Potential Application of iRAP,” in *2010 Electronic (CD) Proceedings of 24th ARRB Conference: Building on 50 years of road and transport research*, Oct. 2010, pp. 18 (p. 161 of abstract book).

Hoque**,** M. M., Anowar, S., and Raihan, M. A., “Towards sustainable road safety in Bangladesh,” in *2008 International Conference on Sustainable Transport for Developing Countries: Concerns, Issues and Options*, Aug. 2008, pp. 63-69.

Hossain, M., and Muromachi, Y., “Understanding crash mechanism on urban expressways using high-resolution traffic data,” *Accident Analysis and Prevention*, Vol. 57, pp. 17–29, 2013.

Hossain, M., *Development of a Real-Time Proactive Road Safety Management System for Urban Expressways*, Ph. D. Thesis, Department of Built Environment, Tokyo Institute of Technology, 2011.

Hossain, M., and Muromachi, Y., "Evaluating location of placement and spacing of detectors for real-time crash prediction on urban expressways," in *2010 Proceedings of the 89th Annual Meeting of Transportation Research Board*, Washington, DC, 2010a.

Hossain, M., and Muromachi, Y., "Development of a real-time crash prediction model for urban expressway," *Journal of EASTS*, Vol. 8, pp. 2091–2107, 2010b.

http://www.ise.bgu.ac.il/faculty/liorr/hbchap15.pdf accessed on June 23, 2013.

http://www.statstutor.ac.uk/resources/uploaded/clusteranalysis.pdf accessed on June 23, 2013.

http://software.ucv.ro/~cmihaescu/ro/teaching/AIR/docs/Lab7-HirarchicalClustering.pdf accessed on June 23, 2013.

http://www.econ.upf.edu/~michael/stanford/maeb7.pdf accessed on June 23, 2013.

http://cran.r-project.org/web/packages/cluster/cluster.pdf accessed on June 23, 2013.

Jain, A. K., Murty, M. N., and Flynn, P.J., "Data Clustering: A Survey," *ACM Computing Surveys*, Vol. 31(3), 1999.

Jang, J.A., Choi, K., and Cho, H., "A fixed sensor-based intersection collision warning system in vulnerable line-of-sight and/or traffic-violation-prone environment," *Intelligent Transportation System*, Vol. 13(4), pp. 1880–1890, 2012.

Kavsek, B., Lavrac, N., and Jovanoski, V., "Apriori-sd: adapting association rule learning to subgroup discovery," *Applied Artificial Intelligence*, Vol. 20, pp. 543–

583, 2006.

Khan, S., Shanmugam, R., Hoeschen, B., “Injury, fatal, and property damage accident models for highway corridors,” *Transportation Research Record*, Vol. 1665, pp. 84–92, 1999.

Kim, K., Nitz, L., Richardson, J., and Li, L., “Personal and Behavioral Predictors of Automobile Crash and Injury Severity,” *Accident Analysis and Prevention*, Vol. 27(4), pp. 469–481, 1995.

King, B., “Step-wise Clustering Procedures,” *Journal of the American Statistical Association*, Vol. 69, pp. 86-101, 1967.

Lee, C., Abdel-Aty, M., and Hsia, L., “Potential real-time indicators of sideswipe crashes on freeways,” *Transportation Research Record*, Vol. 1953, pp. 41–49, 2006.

Lee, C., Hellinga, B., and Saccomanno, F., “Real-time crash prediction model for the application to crash prevention in freeway traffic,” *Transportation Research Record*, Vol. 1840, pp. 67–77, 2003.

Lee, C., Saccomanno, F., and Hellinga, B., “Analysis of crash precursors on instrumented freeways,” *Transportation Research Record*, Vol. 1784, pp. 1–8, 2002.

Liaw, A., and Wiener, M., “Classification and Regression by Random Forest,” *R News*, Vol. 2(3), pp. 18–22, 2002.

Ljubic, P., Todorovski, L., Lavrac, N., and Bullas, J. C., “Time-series analysis of UK traffic accident data,” in *2002 Proceedings of the Fifth International Multi-conference Information Society*, Ljubljana, Svlovenia, 2002, pp. 131–134.

Mahmud, S. M. S., Hoque, M. S., Qazi, A. S., “Road Safety Problems in Bangladesh: Some Major Initiatives, Constraints and Requirements,” *Transport and Communications Bulletin (TCB) for Asia and the Pacific*, Economic and Social

Commission for Asia and the Pacific (ESCAP), No.79, pp. 47-67, 2009.

McCartt, A.T., Northrup, V.S., and Retting, R.A., “Types and characteristics of ramp related motor vehicle crashes on urban interstate roadways in Northern Virginia,” *Journal of Safety Research*, Vol. 35(1), pp. 107–114, 2002.

Miaou, S.-P., and Song, J.J., “Bayesian ranking of sites for engineering safety improvements: decision parameter, treatability concept, statistical criterion and spatial dependence,” *Accident Analysis and Prevention*, Vol. 37(4), pp. 699–720, 2005.

Murtagh, F., “A survey of recent advances in hierarchical clustering algorithms which use cluster centers,” *Comput. J.*, Vol. 26, pp. 354-359, 1984.

Mussone, L., Ferrari, A., and Oneta, M., “An analysis of urban collisions using an artificial intelligence model,” *Accident Analysis and Prevention*, Vol. 31, pp. 705–718, 1999.

Ng, K. S., Hung, W. T., and Wong, W. G., “An algorithm for assessing the risk of traffic accidents,” *Journal of Safety Research*, Vol. 33, pp. 387–410, 2002.

Oh, J., Oh, C., Ritchie, S., and Chang, M., “Real time estimation of accident likelihood for safety enhancement,” *ASCE Journal of Transportation Engineering*, Vol. 131(5), pp. 358–363, 2005a.

Oh, C., Oh, J., Ritchie, S., and Chang, M., “Real time hazardous traffic condition warning system: framework and evaluation,” *IEEE Transactions on Intelligent Transportation Systems*, Vol. 6(3), pp. 265–272, 2005b.

Oh, C., Oh, J., Ritchie, S., and Chang, M., “Real-time estimation of freeway accident likelihood,” in *2001 Proceedings of the 80th Annual Meeting of Transportation Research Board*, Washington, DC, 2001.

Ossiander, E. M., and Cummings, P., “Freeway speed limits and Traffic Fatalities in

Washington State," *Accident Analysis and Prevention*, Vol. 34, pp. 13–18, 2002.

Ossenbruggen, P. J., and Pendharkar, J., "Roadway safety in rural and small urbanized areas," *Accidents Analysis and Prevention*, Vol. 33(4), pp. 485–498, 2001.

Pande, A., and Abdel-Aty, M., "Comprehensive analysis of the relationship between real-time traffic surveillance data and rear-end crashes on freeways," *Transportation Research Record,* Vol. 1953, pp. 31–40, 2006a.

Pande, A., and Abdel-Aty, M., "Assessment of freeway traffic parameters leading to lane-change related collisions," *Accident Analysis and Prevention,* Vol. 38(5), pp. 936–948, 2006b.

Pande, A., and Abdel-Aty, M., "A freeway safety strategy for advanced proactive traffic management," *Journal of Intelligent Transportation Systems*, Vol. 9(3), pp. 145–158, 2005.

Shankar, V., Mannering, F., and Barfield, W., "Statistical Analysis of Accident Severity on Rural Freeways," *Accident Analysis and Prevention*, Vol. 28(3), pp.391–401, 1996.

Shapire, R., Freund, Y., Bartlett, P., and Lee, W., "Boosting the Margin: A New Explanation for the Effectiveness of Voting Methods," *Journal of Annals of Statistics*, Vol. 26(5), pp. 1651-1686, 1998.

Sneath, P., and Sokal, R. (1973) *Numerical Taxonomy*. W. H. Freeman Co., San Francisco, CA.

Sohn, S., and Lee, S., "Data fusion, ensemble and clustering to improve the classification accuracy for the severity of road traffic accidents in Korea," *Safety Science*, Vol. 41(1), pp. 1–14, 2002.

Sohn, S., and Hyungwon, S., "Pattern recognition for a road traffic accident severity in Korea," *Ergonomics*, Vol. 44(1), pp. 101–117, 2001.

Soman, K.P., Diwakar, S., and Ajay, V. (2006) *Insight Into Data Mining: Theory and Practice*. Prentice-Hall, India.

Srisuriyachai, S., “Analysis of road traffic accidents in Nakhon Pathom province of Bangkok using data mining,” Graduate Studies, Bangkok, Mahidol University, 2007.

Statoo Consulting, Switzerland. http://www.statoo.com/en/datamining/ accessed on March 8, 2012.

Strnad, M., Jovic, F., Vorko, A., Kovacic, L., and Toth, D., “Young children injury analysis by the classification entropy method,” *Accident Analysis and Prevention*, Vol. 30, pp. 689–695, 1998.

Ward, J. H., “Hierarchical grouping to optimize an objective function,” *Journal of the American Statistical Association*, Vol. 58, pp. 236-244, 1963.

WHO (2009) *Global status report on road safety, Time for action 2009*. World Health Organization, Department of Violence and Injury Prevention and Disability (VIP), Geneva.

Wong, J., and Chung, Y., "Comparison of Methodology Approach to Identify Causal Factors of Accident Severity," *Transportation Research Record*, Vol. 2083, pp. 190–198, 2008.

Witten, I. H., and Frank, E. (2005) *Data Mining: Practical Machine Learning Tools and Techniques*. 2nd Ed., Morgan Kaufmann, San Francisco, CA.

Xu, C., Wang, W., and Liu, P., “A genetic programming model for real-time crash prediction on freeways,” *IEEE Transactions on Intelligent Transportation Systems*, Vol. 99, pp. 1–13, 2012.

Yang, W. T., Chen, H. C., and Brown, D. B., "Detecting Safer Driving Patterns by a Neural Network Approach," in *1999 Proceedings of Smart Engineering System Design Neural Network*, Evolutionary Programming, Complex Systems and Data Mining, Vol. 9, Nov. 1999, pp. 839–844.

Zeitouni, K., and Chelghoum N., "Spatial decision tree- application to traffic risk analysis," in *2001 proceedings of ACS/IEEE International Conference on Computer Systems and Applications,* Beirut, Lebanon, 2001.

Zelalem, R., "Determining the degree of driver's responsibility for car accident: the case of Addis Ababa traffic office," Addis Ababa University, Addis Ababa, 2009.

Zheng, Z., Ahn, S., and Monsere, C.M., "Impact of traffic oscillations on freeway crash occurrences," *Accident Analysis and Prevention*, Vol. 42(2), pp. 626–636, 2010.

APPENDIX-A

Form No. 34
Bengal Form No. 403Q

1. দুর্ঘটনার ক্রমিক নম্বর

2. প্রাথমিক তথ্য বিবরণী নম্বর

গণপ্রজাতন্ত্রী বাংলাদেশ সরকার
বাংলাদেশ পুলিশ
সড়ক দুর্ঘটনার রিপোর্ট ফরম
Regulation 254 (b)

3. থানা

4. জিলা / মেট্রো পুলিশ

5. দুর্ঘটনা কবলিত গাড়ীর সংখ্যা

6. হতাহত ড্রাইভারের সংখ্যা

7. হতাহত যাত্রীর সংখ্যা

8. হতাহত পথচারীর সংখ্যা

9. দুর্ঘটনার মাত্রা
F. মৃত্যু ঘটিত দুর্ঘটনা
G. মারাত্মক ক্ষত জনিত দুর্ঘটনা
S. সাধারন ক্ষত জনিত দুর্ঘটনা
M. মটর সংঘর্ষ

10. দিন [সোম, মঙ্গল, বুধ, বৃহস্পতি, শুক্র, শনি, রবি]
1 2 3 4 5 6 7

দুর্ঘটনার তারিখ 11. তারিখ 12.মাস 13.বছর
--------- / ---------- / -----------

14. দুর্ঘটনার সময়

রিপোর্ট করার তারিখ

রিপোর্ট করার সময়

15. সংযোগ স্থলের ধরণ
1. সংযোগ স্থল নয়
2.
3.
4.
5.
6. রেলওয়ে ক্রসিং
7. অন্যান্য ---------

16. ট্রাফিক নিয়ন্ত্রণ ব্যবস্থা
1. অনিয়ন্ত্রিত
2. রোড ডিভাইডার দ্বারা নিয়ন্ত্রিত
3. পথচারী পারাপার
4. পুলিশ নিয়ন্ত্রিত
5. ট্রাফিক বাতি নিয়ন্ত্রিত
6. পুলিশ ও ট্রাফিক বাতি নিয়ন্ত্রিত
7. থামা / যেতে দিন সংকেত দ্বারা নিয়ন্ত্রিত
8. অন্যান্য -------

17. সংঘর্ষের ধরন
1. মুখোমুখি
2. পশ্চাদ ভাগ
3. সমকোন
4. পার্শ্ব ঘর্ষণ
5. উল্টে যাওয়া
6. রাস্তার উপরস্থ বস্তুকে আঘাত
7. রাস্তার ধারের বস্তুকে আঘাত
8. দাঁড়ানো গাড়ীকে আঘাত
9. পথচারীকে আঘাত
10. জন্তুকে আঘাত
11. অন্যান্য ----(যেমন গাড়ীর ছাদ থেকে পড়ে যাওয়া)

18. গাড়ী চলাচলের দিক
1. এক মুখী চলাচল
2. উভয়মুখী চলাচল

19. রোড ডিভাইডার
1. আছে
2. নাই

20. আবহাওয়া
1. ভাল
2. বৃষ্টি
3. ঝড়
4. কুয়াশা

21. আলো
1. দিন
2. ভোর / সন্ধ্যা
3. আলোকিত সড়ক (রাতে)
৪. অনালোকিত সড়ক (রাতে)

22. রাস্তার জ্যামিতিক বিবরণ
1. সোজা - সমতল
2. বাঁকানো
3. অসমতল
4. বাঁকা- অসমতল
5. চুড়া

23. রাস্তার উপরিভাগের অবস্থা
1. শুকনা
2. ভেজা
3. কর্দমাক্ত
4. জলমগ্ন /প্লাবিত
5. অন্যান্য ....

24. রাস্তার প্রকারভেদ
1. পাকা
2. ইটের রাস্তা
3. কাঁচা

25. রাস্তার প্রকৃতি
1. ভাল
2. এবড়ো খেবড়ো (রাফ)
3. মেরামত কাজ চলছে

26. রাস্তার শ্রেনী
1. ন্যাশনাল
2. রিজিওনাল
3. ফিডার রোড
4. রুরাল রোড
5. সিটি রোড

27. রাস্তার বৈশিষ্ট্য
1. সাধারণ রাস্তা
2. সেতু
3. কালভার্ট
4. সংকীর্ণ / বাধা প্রাপ্ত
5. স্পীড ব্রেকার

28. এলাকার ধরন
1. শহর
2. গ্রাম এলাকা

অফিসে ব্যবহারের জন্য

29. XY MAP
30. X
31. Y
32. ROUTE
33. KM
34. 100m
35. NODE MAP
36. NODE 1
37. NODE 2

অবস্থান

নগর/শহর/গ্রাম এর নাম..থেকে দুরত্ব :-------------------------------(কি: মি/মি)

রাস্তার নাম : .... মধ্যে
রোড / স্থান (১)................................দুরত্ব : ------------------------------(কি: মি/মি)
রোড / স্থান (২)................................দুরত্ব : ------------------------------(কি: মি/মি)

দ্বিতীয় রাস্তার নাম (শুধু মাত্র সংযোগ স্থানের দুর্ঘটনার ক্ষেত্রে) : ....থেকে দুরত্ব : ------------------------------(কি: মি/মি)

দুর্ঘটনা স্থানের রেখা চিত্র : দুর্ঘটনার স্থান থেকে নিকটবর্তী কিমি পোষ্ট,সেতু বা রাস্তার সংযোগ স্থান বা অন্যান্য যে কোন স্থায়ী গুরুত্বপূর্ন স্থাপনা হইতে দুরত্ব দেখাইয়া চিত্র

সংঘর্ষের রেখা চিত্র : দুর্ঘটনা কবলিত গাড়ী / পথচারী সমুহের চলাচলের দিক এবং অবস্থান সহ সংঘর্ষের পূর্ণ চিত্র

দুর্ঘটনার সংক্ষিপ্ত বিবরণ
....
....
....
....
....
....
....
....
....
....
....
....
....
....

সাক্ষী
১. নাম ও ঠিকানা --------------------------------
----
২. নাম ও ঠিকানা --------------------------------

বিবরণ লিপিবদ্ধকারী অফিসার
নাম/ পদবি ....তারিখ

অনুসন্ধানকারী অফিসার
নাম/ পদবি ....তারিখ

তত্ত্বাবধানকারী অফিসার
নাম/ পদবি.....................................তারিখ

আইনের ধারা

কেসের অবস্থা
1. চার্জসীট
2. ফাইনাল রিপোর্ট
3. তদন্তাধীন

দুর্ঘটনায় ২ এর অধিক যানবাহন, ৬ এর অধিক যাত্রী অথবা ৩ এর অধিক পথচারী হতাহত হইলে অতিরিক্ত ফরমের দরকার হইবে। অতিরিক্ত ফরমে দুর্ঘটনার ক্রমিক নম্বর থানা ও জেলা/মেট্রোপুলিশ এবং দুর্ঘটনার বৎসর উল্লেখ করিয়া এক সাথে গাথিয়া দিতে হইবে।

যানবাহন ১

যানবাহন ১	মালিকের নাম	চালক ১	নাম
মালিকের ঠিকানা		ঠিকানা	
যানবাহন প্রস্তুতকারী	রেজিস্ট্রেশন নম্বর 38. জেলা 39. নম্বর	ড্রাইভিং লাইসেন্স 46. জেলা 47. নম্বর	
40. বৈধ ফিটনেস সার্টিফিকেট 1. আছে 2. নাই 3. প্রযোজ্য নয়	বীমা কৃত 1. ৩য় পার্টি 2. কমপ্রিহেনসিভ	লাইসেন্সের ধরণ এবং যানের শ্রেণী	মেয়াদ অতিক্রান্তের তারিখ

41. যানবাহনের ধরণ
1. বাইসাইকেল
2. রিকসা / ভ্যান
3. ঠেলাগাড়ী
4. মটর সাইকেল
5. বেবী টেক্সী/CNG
6. টেম্পো
7. মাইক্রোবাস
8. মিনি বাস
9. বাস
10. কার
11. জীপ
12. পিক আপ
13. ছোট ট্রাক
14. ভারী ট্রাক
15. আর্টিকুলেটেড ট্রাক
16. ট্যাংকার
17. ট্রাকটর
18. পশু চালিত
19. অন্যান্য ---- (নসিমন/ করিমন ইত্যাদি)

42. যানবাহন চলাচলের ধরণ
1. বামে মোড়
2. ডানে মোড়
3. 'U' মোড়
4. সহযোগ স্থলে আড়াআড়ি অতিক্রম
5. ওভারটেকিং
6. সোজাসুজি চলা
7. পিছনের দিকে চলা
8. হঠাৎ যাত্রা করা
9. হঠাৎ থামা / গতি কমানো
10. পার্ক/দাঁড়ানো অবস্থায়
11. অন্যান্য -------------

48. চালকের লিঙ্গ
1. পুরুষ
2. স্ত্রী

49. চালকের বয়স

50. চালকের ক্ষত
F. মৃত্যু
G. মারাত্মক ক্ষত
S. সাধারণ ক্ষত
N. অক্ষত

43. যানবাহনের মালামাল বোঝাই
1. আইনানুগ
2. বেআইনী/ বিপজ্জনক বোঝাই

44. যানবাহনের ত্রুটি
1. ত্রুটি মুক্ত
2. লাইট
3. ব্রেক
4. ষ্টিয়ারিং
5. টায়ার
6. বহুবিধ
7. অন্যান্য ---------

45. যানবাহনের ক্ষতি (দুর্ঘটনা জনিত)
1. নাই
2. সামনে
3. পিছনে
4. ডানে
5. বামে
6. ছাদে
7. বহুবিধ
8. অন্যান্য----------

51. মদ্যপ কিনা
1. সন্দেহ আছে
2. সন্দেহ মুক্ত

52. সীট বেল্ট / হেলমেট
1. পরিহিত
2. পরিহিত নয়

যানবাহন ২

যানবাহন ২	মালিকের নাম	চালক ২	নাম
মালিকের ঠিকানা		ঠিকানা	
যানবাহন প্রস্তুতকারী	রেজিস্ট্রেশন নম্বর 38. জেলা 39. নম্বর	ড্রাইভিং লাইসেন্স 46. জেলা 47. নম্বর	
40. বৈধ ফিটনেস সার্টিফিকেট 1. আছে 2. নাই 3. প্রযোজ্য নয়	বীমা কৃত 1. ৩য় পার্টি 2. কমপ্রিহেনসিভ	লাইসেন্সের ধরণ এবং যানের শ্রেণী	মেয়াদ অতিক্রান্তের তারিখ

41. যানবাহনের ধরণ
1. বাইসাইকেল
2. রিকসা / ভ্যান
3. ঠেলাগাড়ী
4. মটর সাইকেল
5. বেবী টেক্সী/CNG
6. টেম্পো
7. মাইক্রোবাস
8. মিনি বাস
9. বাস
10. কার
11. জীপ
12. পিক আপ
13. ছোট ট্রাক
14. ভারী ট্রাক
15. আর্টিকুলেটেড ট্রাক
16. ট্যাংকার
17. ট্রাকটর
18. পশু চালিত
19. অন্যান্য ---- (নসিমন/ করিমন ইত্যাদি)

42. যানবাহন চলাচলের ধরণ
1. বামে মোড়
2. ডানে মোড়
3. 'U' মোড়
4. সহযোগ স্থলে আড়াআড়ি অতিক্রম
5. ওভারটেকিং
6. সোজাসুজি চলা
7. পিছনের দিকে চলা
8. হঠাৎ যাত্রা করা
9. হঠাৎ থামা / গতি কমানো
10. পার্ক/দাঁড়ানো অবস্থায়
11. অন্যান্য -------------

48. চালকের লিঙ্গ
1. পুরুষ
2. স্ত্রী

49. চালকের বয়স

50. চালকের ক্ষত
F. মৃত্যু
G. মারাত্মক ক্ষত
S. সাধারণ ক্ষত
N. অক্ষত

43. যানবাহনের মালামাল বোঝাই
1. আইনানুগ
2. বেআইনী/ বিপজ্জনক বোঝাই

44. যানবাহনের ত্রুটি
1. ত্রুটি মুক্ত
2. লাইট
3. ব্রেক
4. ষ্টিয়ারিং
5. টায়ার
6. বহুবিধ
7. অন্যান্য ---------

45. যানবাহনের ক্ষতি (দুর্ঘটনা জনিত)
1. নাই
2. সামনে
3. পিছনে
4. ডানে
5. বামে
6. ছাদে
7. বহুবিধ
8. অন্যান্য----------

51. মদ্যপ কিনা
1. সন্দেহ আছে
2. সন্দেহ মুক্ত

52. সীট বেল্ট / হেলমেট
1. পরিহিত
2. পরিহিত নয়

হতাহত যাত্রীর বিবরণ

একজন যাত্রীর জন্য একটি লাইন পুরণ করুন — * = নীচের বক্স দেখুন

নাম ও ঠিকানা	53. যানবাহন নং	54. লিঙ্গ	55. বয়স	56. * ক্ষত	57. * অবস্থান	58. * কার্যক্রম
1.						
2.						
3.						
4						
5.						
6.						

হতাহত পথচারীর বিবরণ

একজন পথচারীর জন্য একটি লাইন পুরণ করুন — * = নীচের বক্স দেখুন

নাম ও ঠিকানা	59. যানবাহন নং	60. লিঙ্গ	61. বয়স	62. * ক্ষত	63. * অবস্থান	64. * কার্যক্রম
1.						
2.						
3.						

দুর্ঘটনার সহায়ক কারণ

১. মাত্রাতিরিক্ত গতি
২. বেপরোয়া চালান
৩. চালকের ক্লান্তি
৪. সামনের গাড়ির অতি সন্নিকটে চালান
৫. চালকের ভুল সংকেত
৬. ভুল ওভারটেকিং
৭. ভুল ভাবে মোড় নেয়া
৮. মদ্যপ চালক
৯. পথচারীর কার্যক্রম
10. যাত্রীর কার্যক্রম
11. খারাপ রাস্তার জন্য
12. রাস্তার জ্যামিতিক ত্রুটি
13. আবহাওয়া
14. গাড়ীর যান্ত্রিক ত্রুটি
15. বিপজ্জনক বোঝাই
16. টায়ার বার্ষ্ট
17. পশুর কার্যক্রম
18. অন্যান্য ---------- (যেমন: রাস্তার উপর দনাদার/ পিচ্ছিল জিনিস পড়ে থাকা, গতি রোধক, দুর্বল ব্রিজ / কালভার্ট ইত্যাদির কারনে)

65. ☐☐
66. ☐☐
67. ☐☐

*** 56--58 এবং 62--64 এর সহায়ক বক্স**

শুধুমাত্র নমুনার জন্য বৃত্ত দিবেন না	56. যাত্রীর ক্ষত / 62. পথচারীর ক্ষত	57. যাত্রীর অবস্থান	58. যাত্রীর কার্যক্রম	63. পথচারীর অবস্থান	64. পথচারীর কার্যক্রম
	F. মৃত্যু G. মারাত্মক ক্ষত S. সাধারণ ক্ষত	1. গাড়ীর ভিতরে 2. গাড়ীর বাইরে 3. গাড়ীর ছাদে	1. নাই 2. যানে উঠিতেছিল 3. যান হইতে নামিতেছিল 4. যান হইতে পড়িয়া যাওয়া 5. অন্যান্য	1. পথচারি পারাপারে 2. পারাপারের ৫০ মিঃ মধ্যে 3.সড়ক দ্বীপ / ডিভাইডারে 4.রাস্তার উপরে 5.ফুটপাতে 6.রাস্তার পাশে/ সোল্ডারে 7.বাস ষ্টপে	1. নাই 2. রাস্তা পারাপার হওয়া 3. রাস্তার উপর দিয়ে চলা 4. রাস্তার পাশ/ সোল্ডার দিয়ে চলা 5. রাস্তার উপরে খেলা করা

APPENDIX-B

গণপ্রজাতন্ত্রী বাংলাদেশ সরকার
পুলিশ হেডকোয়ার্টার্স, ঢাকা।

প্রজ্ঞাপন

নং-এস, আর, ও   তারিখ ঃ

Police Act, 1861 (V of 1861) এর section 12 এ প্রদত্ত ক্ষমতাবলে মহা-পুলিশ পরিদর্শক, সরকারের পূর্বানুমোদনক্রমে, Police Regulations Bengal, 1943 এর নিম্নরূপ অধিকতর সংশোধন করিল, যথা:-

উপরি-উক্ত Regulations এর Volume II এর B.P. Form No. 34/Bengal Form No-403Q এর পরিবর্তে নিম্নরূপ Form প্রতিস্থাপিত হইবে, যথা:-

B.P. Form No. 34
Bengal Form No. 403Q

1. ACCIDENT REPORT NO.	**BANGLADESH POLICE** Register of Road Traffic Accident (REPORT FORM) [Regulation 254(b)]	3. THANA
2. FIR NO.		4. DISTRICT/MET. POL.

5. NUMBER OF VEHICLES INVOLVED	9. ACCIDENT SEVERITY F. Fatal Accident G. Grievous Accident S. Simple Injury Accident M. Motor Collision	DATE OF OCCURRENCE — 11. DATE 12. MONTH 13. YEAR .... / .... / ....
6. NUMBER OF DRIVER CASUALTIES		14. TIME OF OCCURRENCE
7. NUMBER OF PASSENGER CASUALTIES		Date Of Reporting
8. NUMBER OF PEDESTRIAN CASUALTIES	10. DAY	Time Of Reporting

15. JUNCTION TYPE	16. TRAFFIC CONTROL	17. COLLISION TYPE	18. MOVEMENT / 19. DIVIDER ?
1. Not at Junction 2. + 3. T 4. (staggered) 5. (roundabout) 6. Railway 7. Other	1. No Control 2. Centreline 3. Pedestrian Crossing 4. Police Controlled 5. Traffic Lights 6. Police + Traffic Lights 7. Stop/Give Way sign 8. Other	1. Head On 2. Rear End 3. Right Angle 4. Side Swipe 5. Overturned Vehicle 6. Hit Object in Road 7. Hit Object off Road 8. Hit Parked Vehicle 9. Hit Pedestrian 10. Hit Animal 11. Other	18. MOVEMENT 1. 1-Way Street 2. 2-Way Street 19. DIVIDER ? 1. Yes 2. No

20. WEATHER	21. LIGHT	22. ROAD GEOMETRY	23. SURFACE CONDITION	24. SURFACE TYPE	25. SURFACE QUALITY	26. ROAD CLASS
1. Fair 2. Rain 3. Wind 4. Fog	1. Daylight 2. Dawn/Dusk 3. Night (lit) 4. Night (unlit)	1. Straight + Flat 2. Curve Only 3. Slope Only 4. Curve + Slope 5. Crest	1. Dry 2. Wet 3. Muddy 4. Flooded 5. Other	1. Sealed 2. Brick 3. Earth	1. Good 2. Rough 3. Under Repair	1. National 2. Regional 3. Feeder 4. Rural Road 5. City

27. ROAD FEATURE	28. LOCATION TYPE	OFFICE USE ONLY			
1. None 2. Bridge .... 3. Culvert 4. Narrowing/Restriction 5. Speed Breakers	1. Urban Area 2. Rural Area		29. XY MAP 30. X 31. Y	32. ROUTE 33. KM 34. 100m	35. NODE MAP 36. NODE 1 37. NODE 2

LOCATION

Name of City/Town/Village .... Distance: .... (km/m)

Name of Road .... Between — Landmark 1 .... Distance: .... (km/m)

Landmark 2 .... Distance: .... (km/m)

JUNCTION ACCIDENT ONLY Name of SECOND Road .... Distance: .... (km/m)

LOCATION SKETCH Show site in relation to prominent landmarks such as KM posts, bridges or road intersections. Mark distances to the landmarks	COLLISION DIAGRAM SKETCH mark the position and direction of each vehicle and details of the road layout at the site of the accident

SUMMARY OF ACCIDENT	WITNESSES 1. Name & Address .... 2. Name & Address ....
	RECORDING OFFICER Name/Rank .... Date ....
	INVESTIGATING OFFICER Name/Rank .... Date ....
	SUPERVISING OFFICER Name/Rank .... Date ....
	SECTION OF LAW
	STATUS OF CASE 1. Charge Sheet 2. Final Report 3. Under Investigation

Contd P/2

Additional form(s) will be needed if there are more than 2 vehicles, more than 6 passenger casualties or more than 3 pedestrian casualties
Mark each additional form with the REPORT NUMBER, THANA, DISTRICT/MET POL. and YEAR. Fix forms together

VEHICLE 1

OWNER'S NAME

OWNER'S ADDRESS

VEHICLE MANUFACTURER

VEHICLE REGISTRATION
38. DISTRICT
39. NUMBER

40. VALID FITNESS CERTIFICATE 1. Yes 2. No 3. n/a

INSURANCE COVER 1. Third Party 2. Comprehensive

41. VEHICLE TYPE
1. Bicycle
2. Rickshaw
3. Push Cart
4. Motor Cycle
5. Baby Taxi
6. Tempo
7. Microbus
8. Minibus
9. Bus
10. Car
11. Jeep
12. Pick Up
13. Truck (<3.5t)
14. Heavy Truck
15. Artic. Truck
16. OilTanker
17. Tractor
18. Animal Drawn
19. Other ....

42. VEHICLE MANOEUVRE
1. Left Turn
2. Right Turn
3. 'U' Turn
4. Crossing Road
5. Overtaking
6. Going Ahead
7. Reversing
8. Sudden Start
9. Sudden Stop
10. Parked
11. Other ....

43. VEHICLE LOADING
1. Legal
2. Illegal/Unsafe

44. VEHICLE DEFECT (from MVI report)
1. None
2. Lights
3. Brakes
4. Steering
5. Tyres
6. Multiple
7. Other ....

45. VEHICLE DAMAGE (Sustained in accident)
1. None
2. Front
3. Rear
4. Right
5. Left
6. Roof
7. Multiple
8. Other ....

DRIVER 1

NAME

ADDRESS

DRIVING LICENSE
46. DISTRICT
47. NUMBER

LICENSE TYPE + CATEGORY
EXPIRY DATE

48. DRIVER SEX
1. Male
2. Female

49. DRIVER INJURY
F. Fatal
G. Grievous
S. Simple Injury
N. Not Injured

50. DRIVER AGE

51. ALCOHOL
1. Alcohol Suspected
2. Not Suspected

52. SEAT BELT/HELMET
1. Seat Belt/Helmet Worn
2. Not Worn

VEHICLE 2

OWNER'S NAME

OWNER'S ADDRESS

VEHICLE MANUFACTURER

VEHICLE REGISTRATION
38. DISTRICT
39. NUMBER

40. VALID FITNESS CERTIFICATE 1. Yes 2. No 3. n/a

INSURANCE COVER 1. Third Party 2. Comprehensive

41. VEHICLE TYPE
1. Bicycle
2. Rickshaw
3. Push Cart
4. Motor Cycle
5. Baby Taxi
6. Tempo
7. Microbus
8. Minibus
9. Bus
10. Car
11. Jeep
12. Pick Up
13. Truck (<3.5t)
14. Heavy Truck
15. Artic. Truck
16. OilTanker
17. Tractor
18. Animal Drawn
19. Other ....

42. VEHICLE MANOEUVRE
1. Left Turn
2. Right Turn
3. 'U' Turn
4. Crossing Road
5. Overtaking
6. Going Ahead
7. Reversing
8. Sudden Start
9. Sudden Stop
10. Parked
11. Other ....

43. VEHICLE LOADING
1. Legal
2. Illegal/Unsafe

44. VEHICLE DEFECT (from MVI report)
1. None
2. Lights
3. Brakes
4. Steering
5. Tyres
6. Multiple
7. Other ....

45. VEHICLE DAMAGE (Sustained in accident)
1. None
2. Front
3. Rear
4. Right
5. Left
6. Roof
7. Multiple
8. Other ....

DRIVER 2

NAME

ADDRESS

DRIVING LICENSE
46. DISTRICT
47. NUMBER

LICENSE TYPE + CATEGORY
EXPIRY DATE

48. DRIVER SEX
1. Male
2. Female

49. DRIVER INJURY
F. Fatal
G. Grievous
S. Simple Injury
N. Not Injured

50. DRIVER AGE

51. ALCOHOL
1. Alcohol Suspected
2. Not Suspected

52. SEAT BELT/HELMET
1. Seat Belt/Helmet Worn
2. Not Worn

PASSENGER CASUALTIES

Complete 1 FULL line for each passenger casualty * = See Reference boxes below

NAME AND ADDRESS	53. VEH. NO	54. SEX	55. AGE	56. * INJURY	57. * POSITION	58. * ACTION
1.						
2						
3						
4.						
5						
6.						

PEDESTRIAN CASUALTIES

Complete 1 FULL line for each pedestrian casualty * = See Reference boxes below

NAME AND ADDRESS	59. VEH. NO	60. SEX	61. AGE	62. * INJURY	63. * LOCATION	64. * ACTION
1.						
2						
3.						

FOR REFERENCE ONLY

DO NOT CIRCLE

56. PASSENGER INJURY
62. PEDESTRIAN INJURY
F. Fatal
G. Grievous Injury
S. Simple Injury

57. PASSENGER POSITION
1. Inside Vehicle
2. Outside Vehicle
3. On Roof

58. PASSENGER ACTION
1. No action
2. Boarding
3. De-boarding
4. Falling off
5. Other

63. PEDESTRIAN LOCATION
1. On pedestrian crossing
2. Within 50m of ped.crossing
3. Central island/divider
4. Road centre
5. Footpath
6. Road side
7. Bus stop

64. PEDESTRIAN ACTION
1. No action
2. Crossing the road
3. Walking along the road
4. Walking along road side
5. Playing on the road

CONTRIBUTORY FACTORS
1. Speeding
2. Careless driving
3. Driver fatigue
4. Driving too close
5. Bad driver signals
6. Bad overtaking
7. Bad turning
8. Drunk driver
9. Pedestrian action
10. Passenger action
11. Road condition
12. Road Feature
13. Weather
14. Vehicle Defect
15. Unsafe Loading
16. Tyre Burst
17. Animal Action
18. Other ....

65.
66.
67.

APPENDIX-C

সড়ক দুর্ঘটনার রিপোর্ট ফরম পুরণের নির্দেশিকা

দ্বিতীয় সংস্করণ, জানুয়ারী ২০১০

দুর্ঘটনা রিসার্চ ইন্সটিটিউট (ARI)

বাংলাদেশ প্রকৌশল বিশ্ববিদ্যালয় (BUET)
ঢাকা-১০০০

ভূমিকা

বাংলাদেশ পুলিশের নতুন সড়ক দুর্ঘটনার রিপোর্ট ফরম যথাযথভাবে পুরণের সুবিধার্থে এই নির্দেশিকা প্রকাশ করা হলো। এই পুস্তিকার শেষে একটি পূরণকৃত দুর্ঘটনার রিপোর্ট ফরম দেয়া হলো।

সড়ক দুর্ঘটনার রিপোর্ট ফরমটিতে দুই পৃষ্ঠায় তথ্য লিখার জন্য সর্বমোট ৬৭টি ঘর আছে। এই ঘরসমূহ পুরণের সময় প্রায় ক্ষেত্রেই প্রযোজ্য উত্তরে শুধু গোলদাগ দিতে হবে। অনুসন্ধানকারী অফিসার (Investigating Officer) ফরমটির সম্পূর্ণ অংশ পড়ে, প্রতিটি ঘর ক্রমানুযায়ী যথাযথভাবে পূরণ করবে।

থানা থেকে পূরণকৃত ফরমের অনুলিপি রির্পোটকারী থানায় সংরক্ষণ করতে হবে। মূল ফরমটি পুলিশ সুপার অফিসে পাঠাতে হবে। পুলিশ সুপারগণ পূরণকৃত ফরমসমূহ সংশ্লিষ্ট রেঞ্জ এর এক্সিডেন্ট ডাটা ইউনিটে (ADU) পাঠাবেন। মেট্রোপলিটন এলাকায় থানা থেকে পূরণকৃত ফরম সরাসরি মেট্রোপলিটন পুলিশ কমিশনারের অফিসে অবস্থিত এক্সিডেন্ট ডাটা ইউনিটে (ADU) পাঠাবেন। প্রত্যেক ডিআইজি/মেট্রোপলিটন পুলিশ কমিশনারের অফিসে অবস্থিত এক্সিডেন্ট ডাটা ইউনিট (ADU) দুর্ঘটনার ফরমগুলো থেকে ডাটা MAAP5 Software-এর মাধ্যমে কম্পিউটারে এন্ট্রি করবে। ডিআইজি/মেট্রোপলিটন পুলিশ কমিশনারের দপ্তর থেকে এন্ট্রিকৃত Database CD/Pendrive/E-mail-এর মাধ্যমে ঢাকাস্থ পুলিশ সদর দপ্তরে পাঠাবে। পুলিশ সদর দপ্তর হতে এন্ট্রিকৃত Database CD/Pendrive-এর মাধ্যমে রোড সেফটি সেলে পাঠাতে হবে। রোড সেফটি সেল, জাতীয় সড়ক নিরাপত্তা কাউন্সিলের দায়িত্ব পালনের অংশ হিসাবে তথ্যগুলো সংগ্রহ, বিশ্লেষণ এবং বার্ষিক রিপোর্ট তৈরী করে থাকে এবং এরপর তা' বিভিন্ন সংস্থায় পাঠানো হয়। সড়ক দুর্ঘটনার তথ্য সরকারের নীতি নির্ধারণসহ বিভিন্ন সংস্থার প্রয়োজনে এবং সড়ক দুর্ঘটনা রোধ করার লক্ষ্যে বিভিন্ন গবেষণা প্রতিষ্ঠানের প্রয়োজনে সরবরাহ করা হয়।

অসম্পূর্ণ ও ভুলভাবে পূরণকৃত ফরম সম্পূর্ণ ও শুদ্ধভাবে পূরণ করার জন্য সংশ্লিষ্ট থানায়/অনুসন্ধানকারী কর্মকর্তার নিকট ফেরত পাঠাতে হবে। রিপোর্টকারী থানা তদন্ত নথির জন্য আরও বিস্তারিত মানচিত্র, মৃত্যুর পরবর্তী রিপোর্ট, গাড়ীর পরিদর্শণ রিপোর্ট ইত্যাদির প্রয়োজন হ'তে পারে, তবে এগুলি রিপোর্টকারী থানায় রেখে দিতে হবে।

ফরমটির যেসব ঘরের প্রথমে নম্বর যুক্ত আছে (১ হইতে ৬৭ পর্যন্ত) ঐগুলি কম্পিউটারে সংরক্ষিত হবে। তা' ছাড়াও দুর্ঘটনার লিখিত বিবরণ ও দুর্ঘটনার স্থান কম্পিউটারে সংরক্ষিত থাকবে।

এই রিপোর্ট ফরমটি অনুসন্ধানকারী অফিসার কর্তৃক দুর্ঘটনার স্থানেই অথবা যত তাড়াতাড়ি সম্ভব পূরণ করতে হবে।

(ডঃ মোঃ সামছুল হক)
পরিচালক
দুর্ঘটনা রিসার্চ ইন্সটিটিউট ও
অধ্যাপক, পুরকৌশল বিভাগ
বাংলাদেশ প্রকৌশল বিশ্ববিদ্যালয়

(কিউ.এ.এস.এম.জাকারিয়া ইসলাম)
ডাটা বেইজ স্পেশালিষ্ট
দুর্ঘটনা রিসার্চ ইন্সটিটিউট
বাংলাদেশ প্রকৌশল বিশ্ববিদ্যালয়

তারিখ : জানুয়ারী ২০১০

বি.দ্র. যেসব মেট্রোপলিটন এলাকায় এক্সিডেন্ট ডাটা ইউনিট স্থাপিত হয়নি, সেসব এলাকায় রেঞ্জ অফিসে স্থাপিত ইউনিটই ডাটা এন্ট্রির কাজ করবে।

সড়ক দুর্ঘটনার রিপোর্ট ফরম পূরণ করার পদ্ধতি

(১) দুর্ঘটনার বিস্তারিত বিবরণ ঃ

1| দুর্ঘটনার রিপোর্ট নম্বর ঃ দুর্ঘটনার রিপোর্ট নম্বর রিপোর্টকারী থানা বা আঞ্চলিক হেড কোয়ার্টার কর্তৃক দেয় রিপোর্টের ক্রমিক নম্বর। প্রত্যেক থানা বা আঞ্চলিক অফিস প্রতি বৎসর ০০০১ হতে শুরু করে এই ক্রমিক নম্বর দিবে। প্রতিটি থানা একটি করে সড়ক দুর্ঘটনার হিসাব বই রাখবে যাতে দুর্ঘটনার সময়ানুক্রম পাওয়া যায় ও হেড কোয়ার্টারে ফেরত না দেয়া রিপোর্টের হদিস পাওয়া যায়। এই প্রশিক্ষণ ম্যানুয়েলের শেষে একটি হিসাব বই-এর নমুনা দেয়া হলো। এই দুর্ঘটনার রিপোর্ট নম্বরের সাথে এফ.আই.আর বা এম.সি.আর নম্বর গুলিয়ে ফেলা যাবে না।

2| প্রাথমিক তথ্য বিবরণী নম্বর ঃ থানা কর্তৃক কেস প্রতি দেয়া প্রাথমিক তথ্য বিবরণী (FIR) নম্বর।

3| থানা ঃ দুর্ঘটনার রিপোর্টকারী থানা/পুলিশ ষ্টেশন সমূহের নামের তালিকা প্রতিটি জেলা ও মেট্রোপলিটন পুলিশ বাহিনীতে রক্ষিত আছে।

4| জেলা/মেট্রোপলিটন ঃ পুলিশ জেলা বা মেট্রোপলিটন পুলিশ বাহিনীর নাম।

5| দুর্ঘটনা কবলিত গাড়ির সংখ্যা ঃ দুর্ঘটনা কবলিত সর্বমোট গাড়ির সংখ্যা। এর প্রতিটি গাড়ির জন্য অত্র ফরমের সম্পৃক্ত যানবাহন/চালক অংশ পূরণ করতে হবে।

6| হতাহত চালকের সংখ্যা ঃ দুর্ঘটনায় নিহত বা আহত চালকের মোট সংখ্যা।

7| হতাহত যাত্রীর সংখ্যা ঃ দুর্ঘটনায় নিহত বা আহত যাত্রীর মোট সংখ্যা। এর প্রতি যাত্রীর জন্য অত্র ফরমের সম্পৃক্ত যাত্রীর লাইন/অংশ পূরণ করতে হবে।

8| হতাহত পথচারীর সংখ্যা ঃ দুর্ঘটনায় নিহত বা আহত পথচারীর মোট সংখ্যা। এর প্রতি পথচারীর জন্য অত্র ফরমের সম্পৃক্ত যাত্রীর লাইন/অংশ পূরণ করতে হবে।

9| দুর্ঘটনার মাত্রা ঃ F = মৃত্যুঘটিত দুর্ঘটনা। যেখানে দুর্ঘটনার ৩০ দিনের মধ্যে কোন ব্যক্তি মৃত্যুবরণ করে। G = মারাত্মক ক্ষতজনিত দুর্ঘটনা। যেখানে দুর্ঘটনায় কোন ব্যক্তি মারাত্মকভাবে আহত হয়, তবে কেউ মৃত্যুবরণ করে না। S = সাধারণ ক্ষতজনিত দুর্ঘটনা। যেখানে কোন ব্যক্তি সাধারণভাবে আহত হয়। তবে কেউ মৃত বা মারাত্মকভাবে আহত হয় না। M = মোটর দুর্ঘটনা। যেখানে দুর্ঘটনায় কেউ হতাহত হয় না, কিন্তু গাড়ি বা সম্পদের ক্ষতি সাধিত হয়।

দুর্ঘটনার মাত্রা হতাহতের সংখ্যার উপর নির্ভর করে না বরং হতাহতদের মধ্যে সর্বোচ্চ আঘাতের মাত্রার উপর নির্ভরশীল। যেমন, কোন দুর্ঘটনায় যদি ২০ জন লোক সাধারণভাবে আহত (S) হয় ও ১ জন মারাত্মকভাবে আহত (G) হয় তবে দুর্ঘটনার মাত্রা মারাত্মক ক্ষতজনিত দুর্ঘটনা ধরতে হবে।

10| দিন ঃ সপ্তাহের যে দিন/বারে (সোম,মঙ্গল,বুধ -----) দুর্ঘটনা সংঘটিত হয়।

দুর্ঘটনার তারিখ ঃ

11| তারিখ ঃ মাসের যে তারিখে দুর্ঘটনা সংঘটিত হয়।

12| মাস ঃ যে মাসে দুর্ঘটনা সংঘটিত হয়।

13| বৎসর ঃ যে বৎসর দুর্ঘটনা সংঘটিত হয়।

14| দুর্ঘটনার সময় ঃ দুর্ঘটনা যে সময় সংঘটিত হয়। ২৪ ঘন্টার দিনকে ব্যবহার করতে হবে। উদাহরণ স্বরূপঃ সকাল ৯টা = ০৯.০০, রাত্র ৯টা = ২১.০০। তবে **এ পদ্ধতিতে যদি কোন দুর্ঘটনা রাত ঠিক ১২:০০টায় সংঘটিত হয় তবে**

দুর্ঘটনার সময় ০০:০০ বা ২৪:০০ না লিখে ০০:০১ লিখতে হবে।

রিপোর্ট করার তারিখ ঃ পুলিশের নিকট দুর্ঘটনার রিপোর্ট (FIR) করার দিন, মাস ও বৎসর।

রিপোর্ট করার সময় ঃ পুলিশের নিকট দুর্ঘটনার রিপোর্ট করার সময়।

15| সংযোগ স্থলের ধরণ ঃ দুর্ঘটনার স্থানের ধরণ বুঝে যথাযথ নম্বরে গোল দাগ দিতে হবে। যদি দুর্ঘটনাটি কোন রাস্তার সংযোগস্থলে সংঘটিত হয় তবে এই ফরমের দুর্ঘটনার অবস্থান অংশে দ্বিতীয় সড়কের নাম লিখতে হবে। এছাড়া সংঘর্ষের রেখা চিত্রের ঘরে রাস্তার সংযোগস্থলের যে রেখাচিত্র আঁকা হবে তা এই ঘরের রাস্তার সংযোগ স্থলের ধরনের সাথে অবশ্যই মিল থাকতে হবে।

উল্লেখ্য, দুর্ঘটনাটি সংযোগ স্থলের ২০ মিটারের মধ্যে সংঘটিত হয়ে থাকলে তা' সংযোগ স্থলে হয়েছে ধরে চিহ্নিত করতে হবে।

16| ট্রাফিক নিয়ন্ত্রণ ব্যবস্থা ঃ দুর্ঘটনার স্থানে অবস্থিত যানবাহন নিয়ন্ত্রণ ব্যবস্থার সাথে মিল রেখে যথাযথ ঘরে গোল দাগ দিতে হবে।

17| সংঘর্ষের ধরণ ঃ দুর্ঘটনার সংঘর্ষের ধরণ বুঝে যথাযথ চিহ্নে গোল দাগ দিতে হবে। সংঘর্ষের রেখাচিত্রে এই ঘরের সংঘর্ষের ধরনের দাগের সাথে মিল থাকতে হবে। এটা মনে রাখতে হবে যে, মুখোমুখি, পশ্চাদভাগে, সমকোন ও পার্শ্ব ঘর্ষণ জাতীয় সংঘর্ষের জন্য অন্ততঃ দুটি গাড়ি জড়িত থাকবে। একটি মাত্র গাড়ি কোন বস্তু বা পথচারীকে আঘাত করলে অথবা রাস্তার উপর উল্টে গেলে বা পাশে খাদে পড়ে গেলে এ চারটি ধরণ ব্যবহৃত হবে না।

মুখোমুখি ঃ যখন দু'টি গাড়ি মুখোমুখি সংঘর্ষে নিপতিত হয়।

পশ্চাদভাগঃ যখন একটি গাড়ি আরেকটি গাড়ির পশ্চাদভাগে আঘাত করে।

সমকোন ঃ যখন একটি গাড়ি অন্য গাড়ির পার্শ্বে প্রায় ৯০ ডিগ্রী কোণাকুনি আঘাত করে।

পার্শ্ব ঘর্ষণ ঃ যখন দুটি গাড়ি পরস্পরের পার্শ্ব ঘর্ষণে লিপ্ত হয়। গাড়ি দুটি একই দিকে বা বিপরীত দিকে গতিশীল থাকতে হবে।

18| গাড়ী চলাচলের দিক ঃ দুর্ঘটনাস্থলের রাস্তায় গাড়ি চলাচলের দিক নির্দেশের যথাযথ ঘরে গোল দাগ দিতে হবে।

একমুখি রাস্তা ঃ যখন রাস্তায় গাড়ি শুধু একদিকে চলাচল করে।

উভয়মুখি রাস্তা ঃ যখন রাস্তায় গাড়ি শুধু উভয়দিকেই চলাচল করে।

19| রোড ডিভাইডার ঃ দুর্ঘটনাস্থলের রাস্তার অবস্থা দেখে যথাযথ ঘরে গোল দাগ দিতে হবে।

আছেঃ রাস্তার মাঝ বরাবর কম উচ্চতার দেয়াল (সড়ক দ্বীপ) থাকলে এবং গাড়ি বিপরীত দিকে যেতে না পারলে।

নাই ঃ উপরের অবস্থার বিপরীত।

20| আবহাওয়া ঃ দুর্ঘটনার সময় আবহাওয়ার অবস্থা দেখে যথাযথ ঘরে গোল দাগ দিতে হবে।

21| আলো ঃ দুর্ঘটনার সময় আলোর অবস্থা দেখে যথাযথ ঘরে গোল দাগ দিতে হবে।

22| রাস্তার জ্যামিতিক বিবরণ ঃ দুর্ঘটনার সময় রাস্তার বাস্তব অবস্থা দেখে যথাযথ ঘরে গোল দাগ দিতে হবে।

চুড়া ঃ এটা পাহাড়ের সর্বোচ্চ অবস্থানকে বোঝায় যেখানে উভয় দিক থেকে আগত গাড়িগুলির দৃষ্টিসীমা কমে যায় অর্থাৎ ড্রাইভার সামনে বেশি দুর দেখতে পায়না।

23| রাস্তার উপরিভাগের অবস্থা ঃ দুর্ঘটনাস্থলের রাস্তার উপরিভাগের অবস্থা দেখে যথাযথ ঘরে গোলদাগ দিতে হবে।

24| রাস্তার প্রকারভেদ ঃ দুর্ঘটনা স্থলের রাস্তার উপরিভাগের প্রকারভেদ দেখে যথাযথ ঘরে গোলদাগ দিতে হবে।

25| রাস্তার প্রকৃতি ঃ দুর্ঘটনা স্থলের রাস্তার গুণাগুণ বিচার করে যথাযথ ঘরে গোলদাগ দিতে হবে।

26| রাস্তার শ্রেণী ঃ দুর্ঘটনা স্থলের রাস্তার শ্রেণী বিন্যাস নির্দেশক ঘরে গোলদাগ দিতে হবে। গুরুত্ব নির্বিশেষে প্রধান প্রধান শহরের সকল রাস্তাকে সিটি রোড হিসাবে দেখাতে হবে।

27| রাস্তার বৈশিষ্ট্য ঃ দুর্ঘটনাস্থলের রাস্তার বিশেষ বৈশিষ্ট্য নির্দেশক ঘরে গোলদাগ দিতে হবে।

সাধারণ রাস্তা ঃ যাতে বিশেষ কোন বৈশিষ্ট্য নেই।

সেতু ঃ দুর্ঘটনাটি যদি সেতুর উপর অথবা তার ২০ মিটারের মধ্যে সংঘটিত হয়ে থাকে তবে এই ঘরে গোলদাগ দিতে হবে। দাগের উপর সেতুর / নদীর নাম লিখতে হবে।

কালভার্ট ঃ দুর্ঘটনাটি যদি কোন কালভার্টের উপর অথবা কালভার্টের কারণে হয়ে থাকে তবে এই ঘরে গোলদাগ দিতে হবে।

সংকীর্ণ/বাধাপ্রাপ্ত ঃ দুর্ঘটনা স্থলে যদি কোন অস্থায়ী কারনের (যেমন হাট বাজার/গাড়ী থামানো/রাস্তা মেরামত কাজ ইত্যাদি) জন্য রাস্তা সংকীর্ণ হয়ে গাড়ী চলাচলে বাধাগ্রস্থ হয় তবে এই ঘরে গোলদাগ দিতে হবে।

28| এলাকার ধরণ ঃ দুর্ঘটনাস্থলের ধরণ বিবেচনা করে যথাযথ ঘরে গোল দাগ দিতে হবে।

শহর এলাকা ঃ যেখানে দুর্ঘটনাটি শহর বা নগরের মত বসতিপূর্ণ এলাকায় সংঘটিত হয়ে থাকে। যদি জায়গাটি শহরের সীমানার বাইরেও হয় তবুও বর্ণনাকারী অফিসার তা' শহর এলাকা বিবেচনা করতে পারেন যদি রাস্তার পার্শ্বে জনবসতি থাকে।

গ্রাম এলাকা ঃ যেখানে দুর্ঘটনাটি বসতিপূর্ণ এলাকার বাইরে সংঘটিত হয়ে থাকে। এর মধ্যে রাস্তাটি বন, আবাদী জমি বা ছোট গ্রামের মধ্য দিয়ে যেতে পারে।

(২) দুর্ঘটনার অবস্থানের তথ্য ঃ

দুর্ঘটনার উপযুক্ত অনুসন্ধান করতে হলে দুর্ঘটনা স্থলের অবস্থান-বৈশিষ্ট্য লিখতে হবে। এটা খুবই প্রয়োজনীয়, এই অংশে দুর্ঘটনাস্থলের বিস্তারিত তথ্যটি লিপিবদ্ধ করবেন, যাতে ভবিষ্যতে যে কেউ ঘটনাস্থল খুঁজে বের করতে পারেন। শুধুমাত্র অফিস ব্যবহারের জন্য ৯টি ঘর আছে। এগুলো কম্পিউটারে বিশ্লেষণের জন্য দুর্ঘটনার অবস্থানের বৈশিষ্ট্যসমূহ কোডভুক্ত করা হবে। এই ঘরগুলো পূরণ করা এই অংশের বিস্তারিত তথ্যাদির উপর নির্ভরশীল। অনেক জায়গায় কোন রাস্তা বা বস্তু বা বসতি থেকে দূরত্ব লিখতে হয়। এই দূরত্ব কিলোমিটার বা মিটারে লিখতে হবে। দূরত্ব লিখতে অপ্রয়োজনীয় কিঃ মিঃ অথবা মিঃ কেটে (অর্থাৎ প্রয়োজনীয় মিঃ অথবা কিঃ মিঃ রেখে) লিখতে হবে।

নগর/শহর/গ্রামের নাম ঃ এই ঘরে দুর্ঘটনা স্থলের নগর, শহর বা গ্রামের নাম লিখতে হবে। বসতি কেন্দ্র থেকে এর দূরত্ব লিখতে হবে। দুরত্ব শূণ্য হতে পারে, তখন ঐ ঘরে শূণ্য (০) লিখতে হবে। যদি দুর্ঘটনাস্থল বসতি থেকে অনেক দূর হয়, তা'হলে সবচেয়ে কাছের নগর/শহর/গ্রমের নাম লিখতে হবে। এই বসতি থেকে দূরত্ব ফরমের ঘরে লিখতে হবে।

দুর্ঘটনার অবস্থান ঃ

রাস্তার নাম ঃ এখানে দুর্ঘটনাস্থলের রাস্তার নাম লিখতে হবে। ন্যাশনাল রাস্তা হলে দুই প্রান্তের নগর/শহরের নামসহ একটি আদর্শ নাম

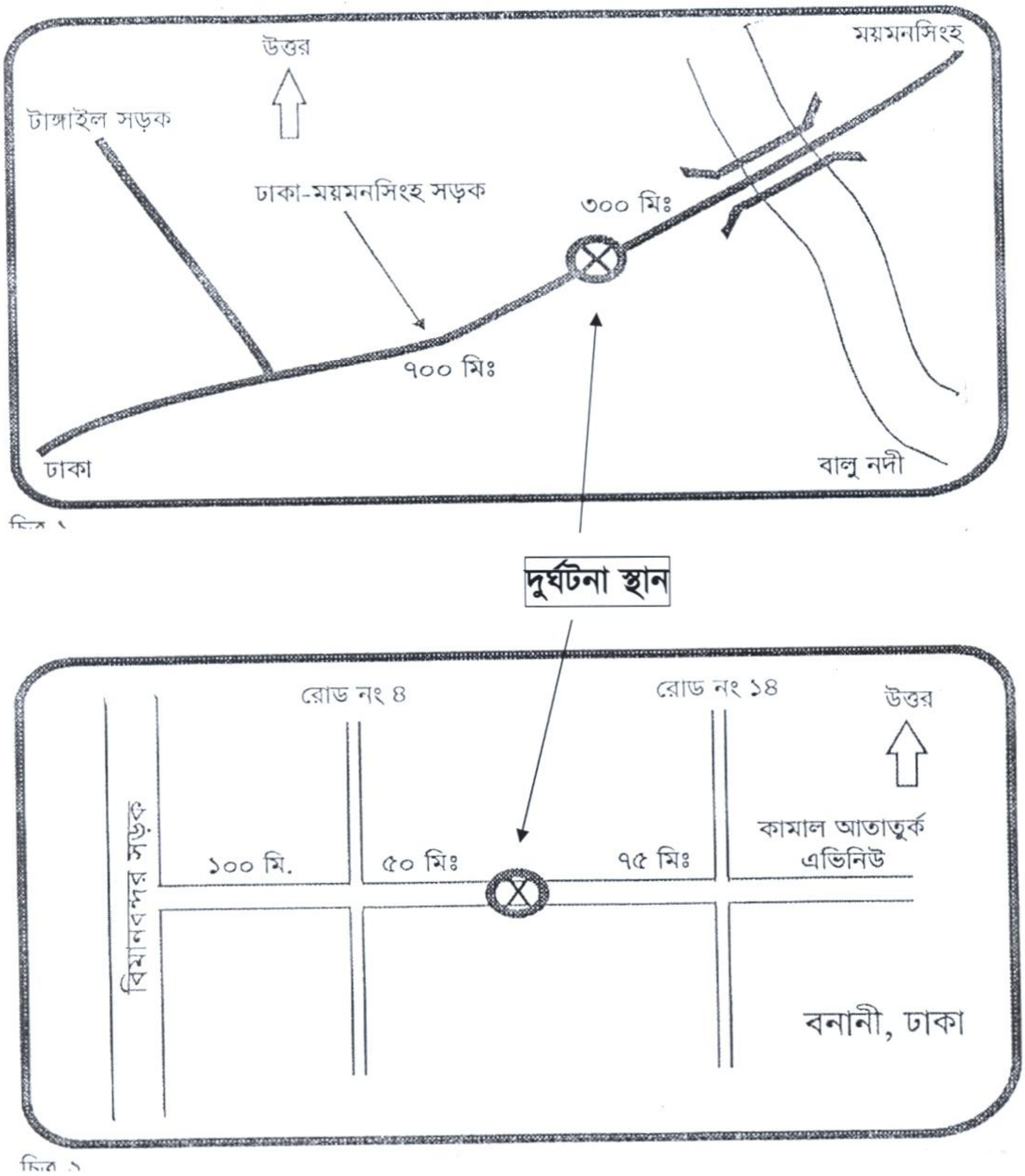
পদ্ধতি ব্যবহার করতে হবে অথবা সড়ক ও জনপথ দপ্তর কর্তৃক ব্যবহৃত সড়ক নম্বর ব্যবহার করতে হবে।

দৃষ্ট বস্তু-১ ঃ এখানে দুর্ঘটনাস্থলের রাস্তার উপর কোন লক্ষণীয় বস্তু / স্থায়ী স্থাপনা যেমন- কিলোমিটার পোষ্ট, সেতু, স্কুল, মাদ্রাসা, মসজিদ, রাস্তার সংযোগ স্থল ইত্যাদির নাম লিখতে হবে। এই লক্ষণীয় বস্তুর/স্থাপনার অবস্থানের দূরত্ব ফরমে জায়গামত লিখতে হবে।

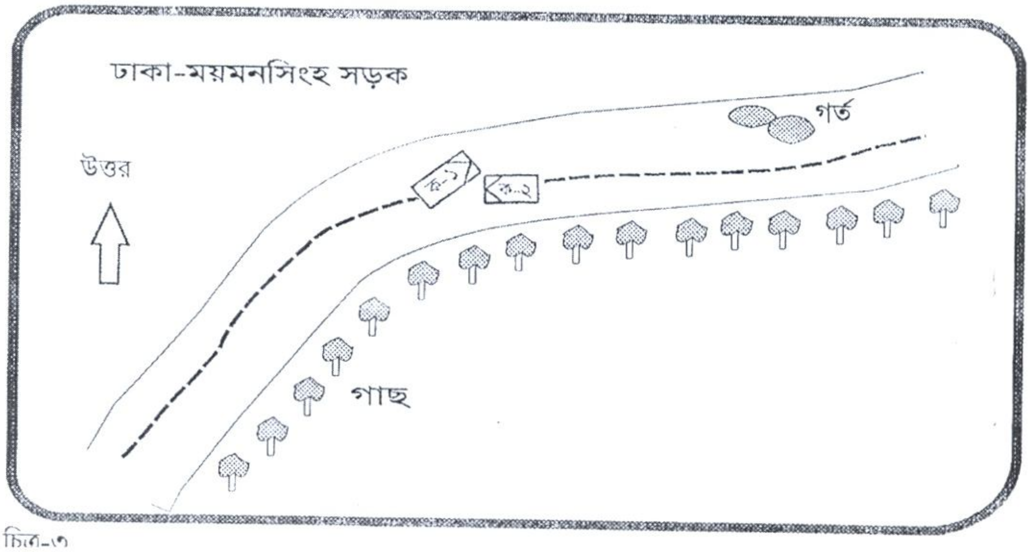
দৃষ্ট বস্তু-২ ঃ এখানে দৃষ্ট বস্তু-১ এর বিপরীত দিকের রাস্তায় অবস্থিত কোন লক্ষণীয় বস্তু / স্থায়ী স্থাপনা যেমন- কিলোমিটার পোষ্ট, সেতু, স্কুল, মাদ্রাসা, মসজিদ, রাস্তার সংযোগ স্থল ইত্যাদির নাম লিখতে হবে। দুর্ঘটনার স্থান থেকে ঐ লক্ষণীয় বস্তুর দূরত্ব ফরমে জায়গামত লিখতে হবে।

শুধুমাত্র সংযোগ স্থানের দুর্ঘটনা ঃ রাস্তার সংযোগ স্থলের দুর্ঘটনার ক্ষেত্রে দুইটি রাস্তারই নাম লিখতে হবে। দুর্ঘটনার স্থান থেকে এই সংযোগ স্থলের দূরত্ব ফরমে জায়গামত লিখতে হবে। দুর্ঘটনাটি যদি এই রাস্তা দুইটির ঠিক সংযোগ স্থলে হয়ে থাকে তবে দূরত্ব শূণ্য লিখতে হবে।

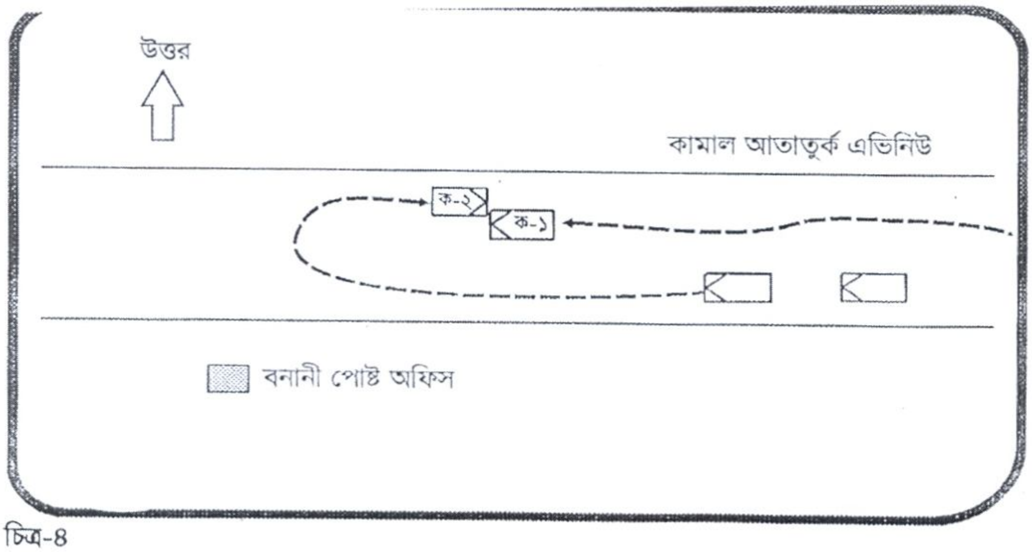
দুর্ঘটনাস্থলের রেখা চিত্র ঃ এই চিত্র অত্যন্ত দরকারী, যাতে ভবিষ্যতে যে কেউই চিত্র দেখে দুর্ঘটনার স্থানটি চিহ্নিত করতে পারে। এখানে শুধুমাত্র রাস্তাটির (বা রাস্তাগুলোর) রেখা চিত্র আঁকলেই চলবে এবং আশে-পাশের দৃষ্ট স্থাপনা সমূহ থেকে দুর্ঘটনার স্থানটির দূরত্ব দেখাতে হবে। মনে রাখতে হবে যে, এই রেখা চিত্রটি শুধুমাত্র দুর্ঘটনাস্থলের অবস্থান জানতে ব্যবহৃত হবে, কাজেই এতে দুর্ঘটনার ধরণের খুঁটিনাটি দেখানোর প্রয়োজন নেই। সংঘর্ষের ধরণের বিবরণ পরে বর্ণিত সংঘর্ষের রেখাচিত্রে দিতে হবে। নিম্নে দুইটি দুর্ঘটনাস্থলের রেখা চিত্রের নমুনা দেয়া হলো।



সংঘর্ষের রেখা চিত্র ঃ এই রেখাচিত্রটি দুর্ঘটনা তদন্তকারীদের জন্য খুবই গুরুত্বপূর্ণ এবং বহু বৎসর পরও তালিকাভুক্ত দুর্ঘটনা-প্রবণ স্থানসমূহের বিশ্লেষণের জন্য প্রয়োজন হয়। এটা একটি সংঘর্ষের রেখা চিত্র মাত্র, আগে বর্ণিত দুর্ঘটনা স্থলের চিত্র নয়। এখানে দুর্ঘটনায় জড়িত প্রত্যেকটি গাড়ির ও পথচারীর রাস্তার উপর অবস্থানস্থল ও চলাচলের পথ দেখাতে হবে। দুর্ঘটনার আগে প্রত্যেকটি গাড়ির গমনপথ ভাঙ্গা দাগ দিয়ে দেখাতে হবে। সংঘর্ষের সময় গাড়িগুলি যে যে দিকে যাচ্ছিল, তা' তীর চিহ্ন দিয়ে দেখাতে হবে। দুর্ঘটনাস্থলের রাস্তার অবস্থানের বিস্তারিত তথ্যাদি সংরক্ষণ করতে হবে। গাড়িগুলোকে ক-১, ক-২ ইত্যাদি প্রতীকে দেখাতে হবে। নিম্নে কয়েকটি সংঘর্ষের রেখাচিত্রের নমুনা দেয়া হল।



চিত্র-৩



চিত্র-৪

(৩) পুলিশের কার্যাদি ঃ

দুর্ঘটনার সংক্ষিপ্ত বিবরণ ঃ এখানে দুর্ঘটনার স্পষ্ট/সঠিক বিবরণ দিতে হবে। গাড়িগুলোকে ক-১, ক-২ ইত্যাদি বলে উল্লেখ করতে হবে। এখানে গাড়ি, পথযাত্রী বা অন্য কিছু, যা দুর্ঘটনার জন্য দায়ী সবই উল্লেখ করতে হবে।

সাক্ষী ঃ এখানে দু'জন সাক্ষীর নাম ও ঠিকানা লিখতে হবে।

বিবরণ লিপিবদ্ধকারী অফিসার ঃ এখানে দুর্ঘটনার বিবরণ লিপিবদ্ধকারী অফিসারের নাম ও পদবী লিখতে হবে।

অনুসন্ধানকারী অফিসার ঃ এখানে দুর্ঘটনার রিপোর্ট ফরম পূরণকারী ও অনুসন্ধানকারী অফিসারের নাম ও পদবী লিখতে হবে।

তত্ত্বাবধানকারী অফিসার ঃ এখানে দুর্ঘটনার রিপোর্ট ফরম পরীক্ষাকারী ও এর সম্পূর্ণতা ও নির্ভুলতা সম্পর্কে অনুমোদনকারী তত্ত্বাবধায়ক অফিসারের নাম ও পদবী লিখতে হবে।

আইনের ধারা ঃ এখানে সড়ক দুর্ঘটনার সংশ্লিষ্ট আইনের ধারা উল্লেখ করতে হবে।

কেসের অবস্থা ঃ তিনটি উল্লেখিত অবস্থার নির্দিষ্ট একটিতে গোল চিহ্ন দিতে হবে।

(৪) যানবাহন/চালক এর বিস্তারিত তথ্য ঃ

দুর্ঘটনা কবলিত প্রতিটি যানবাহনের জন্য এই যানবাহন/চালক অংশ পূরণ করতে হবে। দুর্ঘটনায় ২টির অধিক যানবাহন জড়িত থাকলে অতিরিক্ত ফরম পূরণ করতে হবে ও মূল ফরমের সাথে গেঁথে দিতে হবে। অতিরিক্ত ফরম ব্যবহৃত হলে তাতে দুর্ঘটনার ক্রমিক নং, থানা, জেলা/মেট্রোপুলিশ ও সন লিখতে হবে যাতে তা' সনাক্ত করা যায়। অতিরিক্ত ফরম ব্যবহৃত হলে সবগুলো একসাথে গেঁথে দিতে হবে।

৪.১ যানবাহন এর বিস্তারিত তথ্য ঃ

মালিকের নাম ঃ যানবাহনের মালিকের নাম লিখতে হবে।

মালিকের ঠিকানা ঃ যানবাহনের মালিকের যোগাযোগের ঠিকানা লিখতে হবে।

যানবাহনের প্রস্তুতকারী + তৈরি সন ঃ গাড়িটির বিস্তারিত বিবরণ যথা প্রস্তুতকারী, গঠন প্রকৃতি ও তৈরির সন লিখতে হবে।

38| জেলা ঃ যে জেলায় গাড়িটি রেজিস্ট্রেশন করা হয়েছে। অর্থাৎ ঢাকা, চট্টগ্রাম ইত্যাদি লিখতে হবে।

39| নম্বর ঃ এখানে গাড়িটির কেবলমাত্র রেজিস্ট্রেশন নম্বর লিখতে হবে। এতে গাড়িটির ধরণের সহিত মিল থাকতে হবে।

40| বৈধ ফিটনেস সার্টিফিকেট ঃ প্রযোজ্য ঘরে গোলদাগ দিতে হবে।

আছে ঃ গাড়িটির বৈধ ফিটনেস সার্টিফিকেট আছে।

নাই ঃ গাড়িটির বৈধ বা কোন রকম ফিটনেস সার্টিফিকেট নেই।

প্রযোজ্য নয় ঃ এই ধরণের গাড়ির জন্য ফিটনেস সার্টিফিকেটের প্রয়োজন নেই। (যেমন যন্ত্রবিহীন

গাড়ি এবং নসিমন/করিমন/ভটভটি এই ধরনের স্থানীয়ভাবে তৈরী গাড়ী)।

বীমাকৃত ঃ কৃত বীমার ধরণ বুঝে প্রযোজ্য ঘরে গোল দাগ দিতে হবে।

41| যানবাহনের ধরণ ঃ যানবাহনের ধরণের সাথে মিল রেখে গোল দাগ দিতে হবে। নসিমন/করিমন ধরনের যানবাহনকে অন্যান্য লেখা ঘরে পূরণ করতে হবে।

42| যানবাহন চলাচলের ধরণ ঃ দুর্ঘটনার সময় গাড়িটি যে কৌশলে চলছিল (বা চলার চেষ্টা করছিল) তার সাথে মিল রেখে যথাযথ ঘরে গোল দাগ দিতে হবে। এটা মনে রাখতে হবে যে, পার্ক অবস্থার অর্থ গাড়িটিকে দেখাশুনার কেউ নেই বা গাড়িটি সচল নয়। এতে রাস্তার ভীড়ের/জ্যামের মধ্যে দাঁড়ানো গাড়ি বা রাস্তায় সংযোগ স্থলে পারাপারের সারিবদ্ধ গাড়ি বুঝায় না।

আড়াআড়ি অতিক্রম ঃ এতে গাড়িটি অন্য একটি আড়াআড়ি বড় রাস্তা অতিক্রম করে সম্মুখে যাওয়া বুঝায়।

ওভার টেকিং ঃ যদি গাড়িটি অন্য গাড়িকে অতিক্রম করা অবস্থায় থাকে তবে তাকে অগ্রগমন না বলে ওভার টেকিং বলতে হবে।

43| যানবাহনে মালামাল বোঝাই ঃ গাড়িটিতে মালামাল বোঝাই করার ধরণ দেখে যথাযথ ঘরে গোল দাগ দিতে হবে। যদি অনুসন্ধানকারী অফিসারের মতে মালামাল বোঝাই নিরাপদ ও আইনানুগ হয় তবে প্রথম ঘর চিহ্নিত করতে হবে। কিন্তু যদি মালামাল বোঝাই বিপদজনক ও বে-আইনী হয় তবে দ্বিতীয় ঘর চিহ্নিত করতে হবে। বিপদজনক ও বে-আইনী বলতে অতিরিক্ত মালামাল বহন, ছাদে যাত্রী বহন ইত্যাদি বোঝায়।

44| যানবাহনের ত্রুটি ঃ বিআরটিএ কর্তৃক মটরযানের পরিদর্শন রিপোর্ট দাখিল করার পর এই ঘর পূরণ করতে হবে।

45| যানবাহনের ক্ষতি ঃ দুর্ঘটনার জন্য যানবাহনের যে ক্ষতি হয়েছে তার সাথে মিলিয়ে যথাস্থানে গোল দাগ দিতে হবে। দুর্ঘটনার আগে কোন ক্ষতি থাকলে তা' বিবেচনা করা যাবে না। যদি কোন ক্ষতি দেখা না যায় তা'হলে প্রথম ঘরে দাগ দিতে হবে।

4.2 চালকের বিস্তারিত তথ্য ঃ

নাম ঃ এখানে চালকের নাম লিখতে হবে।

ঠিকানা ঃ এখানে চালকের সঙ্গে যোগাযোগের ঠিকানা লিখতে হবে।

46| জেলা ঃ এখানে চালকের ড্রাইভিং লাইসেন্স যে জেলা হইতে ইস্যু করা হয়েছে তা' লিখতে হবে।

47| নম্বর ঃ এখানে চালকের ড্রাইভিং লাইসেন্স নম্বর লিখতে হবে।

লাইসেন্সের ধরণ ঃ লাইসেন্সের ধরণ ও যানবাহনের শ্রেণী লিখতে হবে।

48| চালকের লিঙ্গ ঃ চালক পুরুষ হলে "১" ও স্ত্রীলোক হলে "২" ঘরে গোল দাগ দিতে হবে।

49| চালকের বয়স ঃ এখানে চালকের বয়স বৎসরে লিখতে হবে।

50| চালকের ক্ষত ঃ নিম্নে বর্ণিত যথাযথ অক্ষরযুক্ত ঘরে গোলদাগ দিতে হবে।

F (মৃত্যু) ঃ দুর্ঘটনায় বা দুর্ঘটনার ৩০ দিনের মধ্যে যদি চালক মৃত্যুবরণ করে।

G (মারাত্মক) ঃ দুর্ঘটনায় যদি চালক মারাত্মক আঘাত প্রাপ্ত হয়।

S (সাধারণ) ঃ দুর্ঘটনায় যদি চালক সাধারণ আঘাত প্রাপ্ত হয়।

N (অক্ষত) ঃ দুর্ঘটনায় যদি চালক আঘাত প্রাপ্ত না হয়।

51| মদ্যপ অবস্থা ঃ এখানে চালক মদ্যপ বা সন্দেহ মুক্ত কিনা লিখতে হবে।

52| সীট বেল্ট/হেলমেট ঃ এখানে চালক সীট বেল্ট বাঁধা অবস্থায় ছিল কিনা এবং দ্বিচক্রযানের ক্ষেত্রে চালক হেলমেট পরিহিত ছিল কিনা লিখতে হবে।

(5) হতাহত যাত্রীর বিবরণ ঃ

দুর্ঘটনায় হতাহত প্রত্যেক যাত্রীর জন্য একটি করে লাইন পূরণ করতে হবে। অক্ষত যাত্রীকে অন্তর্ভুক্ত করা যাবে না।

যদি দুর্ঘটনায় ছয় জনের অধিক হতাহত যাত্রী থাকে তবে অতিরিক্ত ফরম পূরণ করতে হবে। যদি অতিরিক্ত ফরম ব্যবহৃত হয়, তবে তাতে দুর্ঘটনার ক্রমিক নম্বর, থানা, জেলা/মেট্রোপুলিশ ও সন উল্লেখ করতে হবে। অতিরিক্ত ফরম ব্যবহৃত হলে সবগুলো ফরম একসাথে গেঁথে দিতে হবে।

এই অংশ পূরণ করতে ফরমের পথচারীর বিবরণ অংশের পাদটিকার 'বি' নির্দেশ দেখা যেতে পারে। এতে গোল দাগ দিতে হবে না, কারণ যাত্রী সংখ্যা বেশি হতে পারে।

দুর্ঘটনায় নিহত/আহত একজন যাত্রীর জন্য এই ফরমের একটি লাইন পূরণ করতে হবে।

53| যানবাহন নম্বর ঃ যাত্রী যে যানবাহনে ভ্রমণরত ছিলেন সেই নম্বর লিখতে হবে (যেমন ১ নং যানবাহন/ ২ নং যানবাহন বা শুধুমাত্র ১,২ ইত্যাদি)। মনে রাখতে হবে যানবাহনের এই নং গাড়ির নম্বর প্লেট/রেজিস্ট্রেশন নং না।

54। যাত্রীর লিঙ্গ ঃ যাত্রী পুরুষ হলে “১” ও স্ত্রীলোক হলে “২” লিখতে হবে।

55। যাত্রীর বয়স ঃ এখানে যাত্রীর বয়স বৎসরে লিখতে হবে।

56। যাত্রীর ক্ষত ঃ এখানে যাত্রীর ক্ষতের সাথে মিলিয়ে নিচের যে কোন একটি অক্ষর লিখতে হবে।

F (মৃত্যু) ঃ দুর্ঘটনায় বা দুর্ঘটনার ৩০ দিনের মধ্যে যদি যাত্রী মৃত্যুবরণ করে।

G (মারাত্মক) ঃ দুর্ঘটনায় যদি যাত্রী মারাত্মক আঘাত প্রাপ্ত হয়।

S (সাধারণ) ঃ দুর্ঘটনায় যদি যাত্রী সাধারণ আঘাত প্রাপ্ত হয়।

57। যাত্রীর অবস্থান ঃ এই জায়গায় যাত্রীর অবস্থানের উপর বর্ণিত নিচের ছকে দেওয়া নম্বরের সাথে মিলিয়ে শুধু একটি নম্বর লিখতে হবে। কোন নম্বরে গোল দাগ দিতে হবে না। কারণ এটা শুধু নির্দেশিকার জন্য। এখানে যাত্রীর অবস্থান “গাড়ীর বাইরে” বলতে - বাসে উঠাকালীন/বাস বা ট্রাকের ছাদে/ রিক্সা ভ্যান ধরনের উন্মুক্ত যানের আরোহীকে বোঝায়।

58। যাত্রীর কার্যক্রম ঃ এই জায়গায় যাত্রীর কার্যক্রমের উপর বর্ণিত নিচের ছকে দেওয়া নম্বরের সাথে মিলিয়ে শুধু একটি নম্বর লিখতে হবে। কোন নম্বরে গোল দাগ দিতে হবে না। কারণ এটা শুধু নির্দেশিকার জন্য। এখানে যাত্রীর কার্যক্রম বলতে - যাত্রী দুর্ঘটনার সময় কি করছিল তা বোঝায়।

(6) হতাহত পথচারীর বিবরণঃ

দুর্ঘটনায় হতাহত প্রত্যেক পথচারীর জন্য একটি করে লাইন পূরণ করতে হবে। অক্ষত পথচারীকে অন্তর্ভুক্ত করা যাবে না।

যদি দুর্ঘটনায় তিন জনের অধিক হতাহত পথচারী থাকে, তা'হলে অতিরিক্ত ফরম পূরণ করতে হবে। যদি অতিরিক্ত ফরম ব্যবহৃত হয়, তবে তাতে দুর্ঘটনার ক্রমিক নম্বর, থানা, জেলা/মেট্রো-পুলিশ ও সন উল্লেখ করতে হবে যাতে তা' সহজেই সনাক্ত করা যায়। এই অতিরিক্ত ফরমে পথচারী সংখ্যা (পথচারী ৪, পথচারী ৫...) উল্লেখ করতে হবে। অতিরিক্ত ফরম ব্যবহৃত হলে সবগুলি ফরম একসাথে গেঁথে দিতে হবে।

এই অংশ পূরণ করতে নিচের পাদটিকার 'বি' নির্দেশ দেখা যেতে পারে। এতে গোলদাগ দিতে হবে না। কারণ পথচারীর সংখ্যা বেশি হতে পারে।

<u>দুর্ঘটনায় নিহত/আহত একজন পথচারীর জন্য এই ফরমের একটি লাইন পুরণ করতে হবে।</u>

59| যানবাহন নম্বর ঃ যে গাড়ি দ্বারা পথচারী আঘাত প্রাপ্ত হয় সেই গাড়ি নম্বর লিখতে হবে (যেমন গাড়ি নং ক-১, ক-২ অথবা শুধু ১,২)। মনে রাখতে হবে যানবাহনের এই নং গাড়ির নম্বর প্লেট/রেজিস্ট্রেশন নং না।

60| পথচারীর লিঙ্গ ঃ পথচারী পুরুষ হলে "১" ও স্ত্রীলোক হলে "২" লিখতে হবে।

61| পথচারীর বয়স ঃ এখানে পথচারীর বয়স বৎসরে লিখতে হবে।

62| পথচারীর ক্ষত ঃ এখানে পথচারীর ক্ষতের সাথে মিলিয়ে নিচের যে কোন একটি অক্ষর লিখতে হবে।

F (মৃত্যু) ঃ দুর্ঘটনায় বা ৩০ দিনের মধ্যে যদি পথচারী মৃত্যুবরণ করে।

G (মারাত্মক) ঃ দুর্ঘটনায় যদি পথচারী মারাত্মক আঘাত প্রাপ্ত হয়।

S (সাধারণ) ঃ দুর্ঘটনায় যদি পথচারী সাধারণ আঘাত প্রাপ্ত হয়।

63| পথচারীর অবস্থান ঃ এই জায়গায় পথচারীর অবস্থানের উপর বর্ণিত নিচের ছকে দেয়া নম্বরের সাথে মিলিয়ে শুধু একটি নম্বর লিখতে হবে। কোন নম্বরে গোল দাগ দিতে হবে না। কারণ এটা শুধু নির্দেশিকার জন্য।

64| পথচারীর কার্যক্রম ঃ এই জায়গায় পথচারীর কার্যক্রমের উপর বর্ণিত নিচের ছকে দেয়া নম্বরের সাথে মিলিয়ে শুধু একটি নম্বর লিখতে হবে। কোন নম্বরে গোল দাগ দিতে হবে না। কারণ এটা শুধু নির্দেশিকার জন্য।

(7) সম্ভাব্য সহায়ক কারণ ঃ

নিচের দেয়া তিনটি ঘরে দুর্ঘটনার সম্ভাব্য সহায়ক কারণ নির্দেশ করা যেতে পারে। এই তিনটি ঘরের ছকে দেয়া সংখ্যা সমূহ থেকে সম্ভাব্য সহায়ক কারণ নির্দেশক সংখ্যা লিখতে হবে। যদি সহায়ক কারণ তিনটির কম হয়, তবে বাকি ঘরগুলি খালি রাখতে হবে।

65| সহায়ক কারণ ১ ঃ দুর্ঘটনার জন্য গুরুত্বপূর্ণ সহায়ক কারণ নির্দেশক সংখ্যা লিখতে হবে।

66| সহায়ক কারণ ২ ঃ দুর্ঘটনার জন্য দ্বিতীয় গুরুত্বপূর্ণ সহায়ক কারণ নির্দেশক সংখ্যা লিখতে হবে। যদি দ্বিতীয় কোন সহায়ক কারণ না থাকে, তা'হলে এই ঘর খালি রেখে দিতে হবে।

67| সহায়ক কারণ ৩ ঃ দুর্ঘটনার জন্য তৃতীয় গুরুত্বপূর্ণ সহায়ক কারণ নির্দেশক সংখ্যা লিখতে হবে। যদি তৃতীয় কোন সহায়ক কারণ না থাকে, তবে এই ঘর খালি রেখে দিতে হবে।

-- সমাপ্ত --

APPENDIX-D

Table D-1. Year-wise pedestrian accidents vs day of week

Year	Day of Week							Total
	1	2	3	4	5	6	7	
2010	143	136	151	150	123	112	146	961
2009	166	194	188	171	156	182	182	1239
2008	256	249	235	272	278	238	267	1795
2007	304	304	315	314	316	299	334	2186
2006	178	203	218	227	205	201	197	1429
2005	169	171	163	166	174	162	177	1182
2004	244	281	229	229	260	217	215	1675
2003	248	237	233	238	250	211	235	1652
2002	282	265	277	302	252	265	284	1927
2001	206	186	210	181	213	190	169	1355
2000	274	256	255	285	267	249	258	1844
1999	267	233	250	267	255	294	280	1846
1998	232	237	212	282	217	236	237	1653
Total	2969	2952	2936	3084	2966	2856	2981	20744

Notes: 1=Monday, 2=Tuesday, 3=Wednesday, 4=Thursday, 5=Friday, 6=Saturday, 7=Sunday
Source: ARI Accident Database 1998-2010

Table D-2. Year-wise pedestrian accidents vs month of year

Year	Month of Year												Total
	1	2	3	4	5	6	7	8	9	10	11	12	
2010	84	89	113	97	94	68	76	82	71	71	62	54	961
2009	120	147	119	140	115	101	98	91	73	68	85	82	1239
2008	204	175	202	155	165	122	161	148	141	106	121	95	1795
2007	166	157	247	209	168	129	185	172	173	221	186	173	2186
2006	156	125	136	139	134	115	115	115	89	98	96	111	1429
2005	120	109	115	112	116	102	100	82	82	74	92	78	1182
2004	168	138	159	120	129	132	150	142	113	136	139	149	1675
2003	128	139	138	138	154	131	170	157	147	138	135	77	1652
2002	173	176	195	160	160	139	140	163	160	166	137	158	1927
2001	146	113	115	110	138	118	102	122	96	100	103	92	1355
2000	192	155	168	134	167	189	161	128	135	142	144	129	1844
1999	172	158	176	130	155	153	186	174	127	147	122	146	1846
1998	141	132	165	131	142	161	171	122	113	108	133	134	1653
Total	1970	1813	2048	1775	1837	1660	1815	1698	1520	1575	1555	1478	20744

Notes: 1=January, 2=February, 3=March, 4=April, 5=May, 6=June, 7=July, 8=August, 9=September, 10=October, 11=November, 12=December
Source: ARI Accident Database 1998-2010

Table D-3. Year-wise pedestrian accidents vs time of occurrence

Year	Time of Occurrence																									Total
	00-01	01-02	02-03	03-04	04-05	05-06	06-07	07-08	08-09	09-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17	17-18	18-19	19-20	20-21	21-22	22-23	23-24	?*	
2010	25	27	20	21	24	32	28	48	41	52	70	57	56	48	38	42	55	64	41	47	39	25	32	24	5	961
2009	26	23	22	18	24	38	43	58	57	86	79	94	89	66	58	76	74	51	57	63	40	42	26	26	3	1239
2008	65	25	17	32	33	61	62	89	101	111	124	115	115	86	66	82	108	113	69	85	65	56	54	41	20	1795
2007	39	42	48	38	44	65	80	113	123	139	143	175	150	120	101	119	125	128	74	83	71	69	45	35	17	2186
2006	98	20	26	21	34	48	66	70	63	79	104	90	101	77	63	67	84	86	37	50	48	40	25	15	17	1429
2005	117	23	20	14	23	30	34	43	61	72	90	99	71	66	53	61	60	72	32	42	28	24	21	18	8	1182
2004	69	26	21	31	15	49	53	78	81	115	123	135	130	93	73	91	93	102	54	65	54	40	39	27	18	1675
2003	25	21	13	21	26	29	40	69	87	106	128	144	107	97	100	87	125	117	65	71	53	53	33	27	8	1652
2002	20	13	26	26	33	46	57	86	87	137	156	157	134	111	115	121	120	144	68	81	69	47	29	33	11	1927
2001	18	13	15	9	10	28	47	60	65	92	102	106	103	73	77	84	93	106	63	51	32	39	26	22	21	1355
2000	22	17	13	13	21	30	55	82	72	115	155	136	123	122	99	125	147	142	70	83	65	47	42	29	19	1844
1999	19	24	14	13	18	36	55	75	85	114	154	166	130	110	111	112	139	142	76	79	57	40	30	36	11	1846
1998	13	15	14	23	16	24	59	80	109	101	104	134	125	83	102	83	127	114	72	79	59	46	33	29	9	1653
Total	556	289	269	280	321	516	679	951	1032	1319	1532	1608	1434	1152	1056	1150	1350	1381	778	879	680	568	435	362	167	20744

Notes: *"?" means blank data field

Source: ARI Accident Database 1998-2010

Table D-4. Year-wise pedestrian accidents vs junction type

Year	Junction Type								Total
	1	2	3	4	5	6	7	?*	
2010	692	58	88	9	8	0	106	0	961
2009	759	72	109	23	7	0	249	20	1239
2008	1160	134	145	17	13	7	304	15	1795
2007	1352	100	154	24	6	1	478	71	2186
2006	937	97	104	17	10	2	221	41	1429
2005	747	51	103	5	20	2	201	53	1182
2004	1166	65	121	24	32	4	222	41	1675
2003	1190	69	104	16	11	3	212	47	1652
2002	1559	72	129	14	13	6	115	19	1927
2001	1111	49	87	9	6	1	87	5	1355
2000	1513	83	141	16	4	5	80	2	1844
1999	1483	103	166	7	27	3	56	1	1846
1998	1272	100	148	10	31	4	67	21	1653
Total	14941	1053	1599	191	188	38	2398	336	20744

Notes: *"?" means blank data field
1=Not at junction, 2=Cross junction, 3=Tee junction, 4=Staggered tee junction, 5=Roundabouts, 6= Railway/level crossing, 7=Other
Source: ARI Accident Database 1998-2010

Table D-5. Year-wise pedestrian accidents vs traffic control system

Year	Traffic Control									Total
	1	2	3	4	5	6	7	8	?*	
2010	751	10	47	72	3	10	5	63	0	961
2009	851	19	81	135	9	7	0	129	8	1239
2008	1271	30	125	138	5	7	1	183	35	1795
2007	1493	37	143	172	7	16	0	280	38	2186
2006	1024	30	63	157	1	6	4	126	18	1429
2005	864	29	57	65	7	42	3	89	26	1182
2004	1291	15	65	85	7	53	1	141	17	1675
2003	1270	27	88	83	6	24	1	125	28	1652
2002	1578	20	56	103	13	55	1	92	9	1927
2001	1147	19	26	73	22	22	2	42	2	1355
2000	1606	29	29	60	12	50	2	49	7	1844
1999	1524	40	56	63	3	114	2	42	2	1846
1998	1327	41	38	59	10	106	2	57	13	1653
Total	15997	346	874	1265	105	512	24	1418	203	20744

Notes: *"?" means blank data field

1=No control, 2=Centerline marking, 3=Pedestrian crossing, 4=Police controlled, 5=Traffic lights, 6=Police + Traffic lights, 7=Stop/Give way sign, 8=Other

Source: ARI Accident Database 1998-2010

Table D-6. Year-wise pedestrian accidents vs traffic collision type

Year	Collision Type												Total
	1	2	3	4	5	6	7	8	9	10	11	?*	
2010	12	20	1	5	5	0	2	5	895	0	16	0	961
2009	22	27	2	20	2	3	9	9	1114	0	27	4	1239
2008	29	46	3	27	13	5	7	13	1606	0	42	4	1795
2007	37	45	2	39	27	13	13	7	1947	1	53	2	2186
2006	40	35	0	37	16	8	13	5	1251	0	23	1	1429
2005	21	28	5	27	16	6	7	5	1038	0	23	6	1182
2004	28	40	3	31	14	9	14	4	1484	0	47	1	1675
2003	19	48	1	27	13	6	18	7	1481	0	30	2	1652
2002	24	31	3	23	13	6	17	12	1775	0	23	0	1927
2001	22	48	1	17	14	6	12	11	1206	0	17	1	1355
2000	31	19	1	6	12	7	7	13	1738	1	8	1	1844
1999	24	16	2	10	22	5	15	12	1729	0	9	2	1846
1998	25	30	5	27	13	9	7	5	1509	0	18	5	1653
Total	334	433	29	296	180	83	141	108	18773	2	336	29	20744

Notes: *"?" means blank data field

1=Head on, 2=Rear end, 3=Right angle, 4=Side swipe, 5=Overturn, 6=Hit object in road, 7=Hit object off road, 8=Hit parked vehicle, 9=Hit pedestrian, 10=Hit animal, 11=Other

Source: ARI Accident Database 1998-2010

Table D-7. Year-wise pedestrian accidents vs traffic movement

Year	Traffic Movement			Total
	1	2	?*	
2010	115	846	0	961
2009	292	938	9	1239
2008	456	1329	10	1795
2007	739	1417	30	2186
2006	457	958	14	1429
2005	158	1018	6	1182
2004	159	1506	10	1675
2003	192	1449	11	1652
2002	151	1775	1	1927
2001	74	1279	2	1355
2000	43	1799	2	1844
1999	52	1793	1	1846
1998	71	1579	3	1653
Total	2959	17686	99	20744

Notes: *"?" means blank data field
1=1-Way street, 2=2-Way street
Source: ARI Accident Database 1998-2010

Table D-8. Year-wise pedestrian accidents vs presence of divider in roads

Year	Presence of Divider			Total
	1	2	?*	
2010	225	736	0	961
2009	282	938	19	1239
2008	428	1355	12	1795
2007	486	1637	63	2186
2006	348	1055	26	1429
2005	236	911	35	1182
2004	383	1246	46	1675
2003	292	1313	47	1652
2002	385	1542	0	1927
2001	240	1109	6	1355
2000	386	1456	2	1844
1999	328	1513	5	1846
1998	419	1228	6	1653
Total	4438	16039	267	20744

Notes: *"?" means blank data field
1=Yes, 2=No
Source: ARI Accident Database 1998-2010

Table D-9. Year-wise pedestrian accidents vs weather condition

Year	Weather Condition					Total
	1	2	3	4	?*	
2010	925	17	1	18	0	961
2009	1178	30	2	26	3	1239
2008	1717	47	3	27	1	1795
2007	2083	73	1	21	8	2186
2006	1383	23	4	15	4	1429
2005	1122	34	3	16	7	1182
2004	1616	41	2	14	2	1675
2003	1580	45	4	22	1	1652
2002	1861	45	1	20	0	1927
2001	1303	36	2	14	0	1355
2000	1767	53	3	21	0	1844
1999	1771	60	5	10	0	1846
1998	1598	40	1	9	5	1653
Total	19904	544	32	233	31	20744

Notes: *“?” means blank data field
1=Fair, 2=Rain, 3=Wind, 4=Fog
Source: ARI Accident Database 1998-2010

Table D-10. Year-wise pedestrian accidents vs light condition

Year	Light Condition					Total
	1	2	3	4	?*	
2010	611	117	129	103	1	961
2009	805	169	144	116	5	1239
2008	1221	240	196	132	6	1795
2007	1581	228	207	156	14	2186
2006	1055	156	154	60	4	1429
2005	837	137	117	78	13	1182
2004	1203	200	160	105	7	1675
2003	1186	188	153	120	5	1652
2002	1390	229	182	124	2	1927
2001	983	168	104	99	1	1355
2000	1342	210	153	137	2	1844
1999	1361	224	147	113	1	1846
1998	1190	203	150	103	7	1653
Total	14765	2469	1996	1446	68	20744

Notes: *“?” means blank data field
1=Daylight, 2=Dawn/Dusk, 3=Night (lit), 4= Night (unlit)
Source: ARI Accident Database 1998-2010

Table D-11. Year-wise pedestrian accidents vs road geometric condition

Year	Road Geometric Condition						Total
	1	2	3	4	5	?*	
2010	883	43	18	16	1	0	961
2009	1137	52	22	18	3	7	1239
2008	1637	87	36	21	10	4	1795
2007	2038	68	27	19	6	28	2186
2006	1356	36	18	11	2	6	1429
2005	1084	57	18	7	1	15	1182
2004	1582	68	10	5	2	8	1675
2003	1543	74	12	14	4	5	1652
2002	1819	70	22	13	3	0	1927
2001	1265	57	16	11	4	2	1355
2000	1702	89	29	19	5	0	1844
1999	1726	68	37	11	3	1	1846
1998	1555	61	18	9	5	5	1653
Total	19327	830	283	174	49	81	20744

Notes: *"?" means blank data field
1=Straight + Flat, 2=Curve only, 3=Slope only, 4=Curve + Slope, 5=Crest
Source: ARI Accident Database 1998-2010

Table D-12. Year-wise pedestrian accidents vs road surface condition

Year	Road Surface Condition						Total
	1	2	3	4	5	?*	
2010	919	25	2	0	15	0	961
2009	1177	30	3	1	20	8	1239
2008	1700	64	7	2	20	2	1795
2007	2089	63	4	0	9	21	2186
2006	1391	22	1	1	8	6	1429
2005	1118	38	1	2	7	16	1182
2004	1619	43	0	0	9	4	1675
2003	1585	53	2	0	8	4	1652
2002	1866	52	4	0	5	0	1927
2001	1306	38	0	0	9	2	1355
2000	1762	68	7	2	5	0	1844
1999	1782	58	2	0	4	0	1846
1998	1594	49	3	0	1	6	1653
Total	19908	603	36	8	120	69	20744

Notes: *"?" means blank data field
1=Dry, 2=Wet, 3=Muddy, 4=Flooded, 5=Other
Source: ARI Accident Database 1998-2010

Table D-13. Year-wise pedestrian accidents vs surface type

Year	Surface type				Total
	1	2	3	?*	
2010	939	9	13	0	961
2009	1214	6	16	3	1239
2008	1749	16	27	3	1795
2007	2122	18	27	19	2186
2006	1379	18	26	6	1429
2005	1140	17	15	10	1182
2004	1634	16	20	5	1675
2003	1606	17	26	3	1652
2002	1874	27	26	0	1927
2001	1322	15	16	2	1355
2000	1763	35	46	0	1844
1999	1793	29	24	0	1846
1998	1605	21	22	5	1653
Total	20140	244	304	56	20744

Notes: *"?" means blank data field
1=Sealed, 2=Brick, 3=Earth
Source: ARI Accident Database 1998-2010

Table D-14. Year-wise pedestrian accidents vs surface quality

Year	Surface Quality				Total
	1	2	3	?*	
2010	925	30	6	0	961
2009	1177	43	14	5	1239
2008	1700	70	16	9	1795
2007	2102	50	14	20	2186
2006	1361	46	15	7	1429
2005	1127	35	8	12	1182
2004	1581	53	32	9	1675
2003	1581	45	17	9	1652
2002	1851	51	25	0	1927
2001	1306	38	9	2	1355
2000	1746	72	26	0	1844
1999	1778	48	19	1	1846
1998	1600	32	14	7	1653
Total	19835	613	215	81	20744

Notes: *"?" means blank data field
1=Good, 2=Rough, 3=Under repair
Source: ARI Accident Database 1998-2010

Table D-15. Year-wise pedestrian accidents vs road class

Year	Road Class						Total
	1	2	3	4	5	?*	
2010	389	118	108	118	226	2	961
2009	503	176	155	123	276	6	1239
2008	825	215	196	195	359	5	1795
2007	926	279	389	187	403	2	2186
2006	585	163	216	123	334	8	1429
2005	421	195	128	122	313	3	1182
2004	684	188	224	111	463	5	1675
2003	618	240	250	143	396	5	1652
2002	676	237	347	121	545	1	1927
2001	516	193	242	89	314	1	1355
2000	657	238	328	140	477	4	1844
1999	677	241	287	116	524	1	1846
1998	518	226	168	101	629	11	1653
Total	7995	2709	3038	1689	5259	54	20744

Notes: *"?" means blank data field
1=National, 2=Regional, 3=Feeder, 4=Rural road, 5=City
Source: ARI Accident Database 1998-2010

Table D-16. Year-wise pedestrian accidents vs road feature

Year	Road Feature						Total
	1	2	3	4	5	?*	
2010	930	18	2	6	3	2	961
2009	1175	20	8	15	11	10	1239
2008	1714	25	8	13	9	26	1795
2007	2095	16	8	11	7	49	2186
2006	1361	26	11	9	2	20	1429
2005	1119	16	8	12	2	25	1182
2004	1617	22	6	9	0	21	1675
2003	1596	25	6	9	3	13	1652
2002	1857	31	9	21	5	4	1927
2001	1314	19	8	9	3	2	1355
2000	1793	18	8	20	2	3	1844
1999	1795	21	6	17	4	3	1846
1998	1584	25	8	15	2	19	1653
Total	19950	282	96	166	53	197	20744

Notes: *"?" means blank data field
1=None, 2=Bridge, 3=Culvert, 4=Narrowing/Restriction, 5=Speed breakers
Source: ARI Accident Database 1998-2010

Table D-17. Year-wise pedestrian accidents vs road location

Year	Location			Total
	1	2	?*	
2010	497	456	8	961
2009	614	609	16	1239
2008	742	1020	33	1795
2007	845	1291	50	2186
2006	571	828	30	1429
2005	396	761	25	1182
2004	593	1051	31	1675
2003	558	1074	20	1652
2002	688	1229	10	1927
2001	472	874	9	1355
2000	699	1138	7	1844
1999	707	1134	5	1846
1998	765	871	17	1653
Total	8147	12336	261	20744

Notes: *"?" means blank data field
1=Urban area, 2=Rural area
Source: ARI Accident Database 1998-2010

APPENDIX-E

Table E-1. Year-wise double vehicle accidents vs day of week

Year	Day of Week								?*
	1	2	3	4	5	6	7	Total	
2010	101	96	118	104	111	91	90	711	0
2009	138	140	134	136	128	141	118	935	0
2008	187	173	201	212	173	185	154	1285	1
2007	168	172	191	194	184	169	170	1248	0
2006	128	110	122	127	162	132	134	915	0
2005	102	120	116	126	117	95	134	810	0
2004	156	167	158	166	193	168	155	1163	0
2003	188	213	184	187	190	183	171	1316	0
2002	204	180	200	229	199	214	167	1393	0
2001	133	131	118	136	135	139	138	930	0
2000	220	222	186	242	199	189	218	1476	0
1999	205	188	205	218	205	226	210	1457	0
1998	202	182	200	197	197	196	184	1358	0
Total	2132	2094	2133	2274	2193	2128	2043	14997	1

Notes: *"?" means blank data field
1=Monday, 2=Tuesday, 3=Wednesday, 4=Thursday, 5=Friday, 6=Saturday, 7=Sunday
Source: ARI Accident Database 1998-2010

Table E-2. Year-wise double vehicle accidents vs month of year

Year	Month of Year												Total
	1	2	3	4	5	6	7	8	9	10	11	12	
2010	60	67	63	56	69	74	52	64	46	57	52	51	711
2009	114	96	69	89	80	101	80	81	68	36	50	71	935
2008	136	103	116	131	123	103	91	110	131	84	79	79	1286
2007	98	84	128	108	109	94	116	87	95	119	102	108	1248
2006	81	72	91	85	101	91	82	77	57	63	50	65	915
2005	84	75	78	70	83	90	59	60	56	40	65	50	810
2004	128	102	95	86	103	98	112	83	81	105	86	84	1163
2003	98	94	110	101	120	112	144	114	106	126	119	72	1316
2002	136	108	124	120	121	93	126	116	92	131	95	131	1393
2001	122	58	101	68	97	69	73	64	61	58	74	85	930
2000	149	132	143	125	125	131	123	101	117	108	127	95	1476
1999	140	123	131	122	137	114	120	115	115	122	106	112	1457
1998	121	131	144	128	121	120	99	121	94	85	87	107	1358
Total	1467	1245	1393	1289	1389	1290	1277	1193	1119	1134	1092	1110	14998

Notes: 1=January, 2=February, 3=March, 4=April, 5=May, 6=June, 7=July, 8=August, 9=September, 10=October, 11=November, 12=December
Source: ARI Accident Database 1998-2010

Table E-3. Year-wise double vehicle accidents vs time of occurrence

Year	Time of Occurrence																									Total
	00-01	01-02	02-03	03-04	04-05	05-06	06-07	07-08	08-09	09-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17	17-18	18-19	19-20	20-21	21-22	22-23	23-24	?*	
2010	10	10	15	10	12	21	29	44	40	41	46	42	66	40	35	46	36	30	35	27	29	17	14	14	2	711
2009	17	14	10	13	16	24	38	51	51	49	60	61	52	46	68	68	54	54	39	42	29	28	27	22	2	935
2008	23	33	16	25	22	40	39	47	62	82	97	82	84	72	66	71	88	70	51	56	62	40	25	29	4	1286
2007	34	24	23	26	32	51	45	53	63	63	93	101	96	73	62	74	70	61	39	51	30	32	20	26	6	1248
2006	74	14	16	23	23	17	30	49	40	75	59	71	54	45	43	57	46	47	16	31	30	21	19	11	4	915
2005	92	17	12	17	19	22	22	41	39	42	48	56	44	37	46	35	35	26	27	32	29	23	17	25	7	810
2004	63	23	12	22	22	30	45	54	55	71	88	92	86	50	56	54	67	59	42	47	39	37	21	25	3	1163
2003	35	20	19	25	25	42	49	55	55	58	91	94	93	76	68	87	83	85	54	50	42	47	32	28	3	1316
2002	20	31	12	13	24	31	49	68	59	82	102	106	116	82	86	76	97	65	54	53	44	46	38	32	7	1393
2001	14	13	10	11	12	23	30	49	45	63	76	63	71	46	63	60	60	48	27	26	30	30	23	34	3	930
2000	32	39	22	14	20	40	43	64	83	88	104	115	83	87	76	91	85	77	52	64	66	50	41	34	6	1476
1999	21	21	27	20	23	36	53	64	58	77	111	93	100	99	95	101	72	86	40	74	55	46	41	40	4	1457
1998	23	21	17	13	18	24	58	68	68	86	99	98	84	71	71	94	93	62	55	58	60	39	42	35	1	1358
Total	458	280	211	232	268	401	530	707	718	877	1074	1074	1029	824	835	914	886	770	531	611	545	456	360	355	52	14998

Notes: *"?" means blank data field
Source: ARI Accident Database 1998-2010

Table E-4. Year-wise double vehicle accidents vs junction type

Year	Junction Type								Total
	1	2	3	4	5	6	7	?*	
2010	473	58	72	11	8	0	88	1	711
2009	534	70	88	14	5	2	214	8	935
2008	806	100	104	14	6	1	240	15	1286
2007	741	71	103	28	8	3	237	57	1248
2006	594	62	88	15	9	1	122	24	915
2005	513	45	79	10	28	1	107	27	810
2004	780	49	119	12	31	5	140	27	1163
2003	942	58	116	17	24	5	127	27	1316
2002	1030	67	146	20	51	10	61	8	1393
2001	729	41	83	10	11	4	49	3	930
2000	1133	113	156	19	6	6	42	1	1476
1999	1109	106	140	10	61	4	24	3	1457
1998	946	129	167	9	64	6	25	12	1358
Total	10330	969	1461	189	312	48	1476	213	14998

Notes: *"?" means blank data field
1=Not at junction, 2=Cross junction, 3=Tee junction, 4=Staggered tee junction, 5=Roundabouts, 6= Railway/level crossing, 7=Other
Source: ARI Accident Database 1998-2010

Table E-5. Year-wise double vehicle accidents vs traffic control system

Year	Traffic Control									Total
	1	2	3	4	5	6	7	8	?*	
2010	563	13	4	65	5	5	0	56	0	711
2009	706	17	5	70	7	5	3	113	9	935
2008	976	30	8	91	4	12	3	142	20	1286
2007	900	42	10	119	7	4	1	146	19	1248
2006	683	27	9	104	0	4	3	74	11	915
2005	603	22	6	51	5	39	0	66	18	810
2004	913	25	6	61	7	64	0	71	16	1163
2003	1018	53	13	74	8	51	3	78	18	1316
2002	1099	32	8	80	12	97	1	58	6	1393
2001	764	28	5	54	19	27	1	29	3	930
2000	1218	38	7	83	11	81	3	34	1	1476
1999	1155	45	4	47	4	171	3	24	4	1457
1998	983	52	10	77	7	185	0	37	7	1358
Total	11581	424	95	976	96	745	21	928	132	14998

Notes: *"?" means blank data field
1=No control, 2=Centerline marking, 3=Pedestrian crossing, 4=Police controlled, 5=Traffic lights, 6=Police + Traffic lights, 7=Stop/Give way sign, 8=Other
Source: ARI Accident Database 1998-2010

Table E-6. Year-wise double vehicle accidents vs traffic collision type

Year	Collision Type												Total
	1	2	3	4	5	6	7	8	9	10	11	?*	
2010	327	254	13	72	4	0	3	34	0	0	4	0	711
2009	381	353	19	101	6	6	0	61	0	0	5	3	935
2008	503	446	23	184	11	11	6	78	0	0	23	1	1286
2007	524	419	19	177	20	10	9	49	0	1	19	1	1248
2006	344	342	20	142	10	5	3	40	0	0	8	1	915
2005	315	268	30	118	10	5	4	46	0	0	11	3	810
2004	457	411	25	153	7	10	7	62	0	0	29	2	1163
2003	515	453	28	178	8	7	3	101	0	0	21	2	1316
2002	451	551	57	218	14	6	2	76	0	0	17	1	1393
2001	323	383	42	105	3	3	1	62	0	0	8	0	930
2000	499	655	56	180	16	3	2	59	0	0	6	0	1476
1999	482	625	23	246	13	0	3	53	0	0	12	0	1457
1998	428	563	58	228	12	5	3	46	0	0	12	3	1358
Total	5549	5723	413	2102	134	71	46	767	0	1	175	17	14998

Notes: *"?" means blank data field

1=Head on, 2=Rear end, 3=Right angle, 4=Side swipe, 5=Overturn, 6=Hit object in road, 7=Hit object off road, 8=Hit parked vehicle, 9=Hit pedestrian, 10=Hit animal, 11=Other

Source: ARI Accident Database 1998-2010

Table E-7. Year-wise double vehicle accidents vs traffic movement

Year	Traffic Movement			Total
	1	2	?*	
2010	79	631	1	711
2009	161	768	6	935
2008	281	995	10	1286
2007	396	837	15	1248
2006	249	658	8	915
2005	101	704	5	810
2004	108	1052	3	1163
2003	129	1183	4	1316
2002	87	1305	1	1393
2001	51	879	0	930
2000	67	1408	1	1476
1999	34	1423	0	1457
1998	30	1326	2	1358
Total	1773	13169	56	14998

Notes: *"?" means blank data field
1=1-Way street, 2=2-Way street
Source: ARI Accident Database 1998-2010

Table E-8. Year-wise double vehicle accidents vs presence of divider in roads

Year	Presence of Divider			Total
	1	2	?*	
2010	141	568	2	711
2009	173	747	15	935
2008	248	1020	18	1286
2007	300	898	50	1248
2006	239	647	29	915
2005	216	569	25	810
2004	280	842	41	1163
2003	252	1022	42	1316
2002	372	1020	1	1393
2001	214	715	1	930
2000	405	1070	1	1476
1999	334	1118	5	1457
1998	436	916	6	1358
Total	3610	11152	236	14998

Notes: *"?" means blank data field
1=Yes, 2=No
Source: ARI Accident Database 1998-2010

Table E-9. Year-wise double vehicle accidents vs weather condition

Year	Weather Condition					Total
	1	2	3	4	?*	
2010	676	19	3	13	0	711
2009	849	35	4	44	3	935
2008	1197	57	1	31	0	1286
2007	1156	66	2	21	3	1248
2006	860	23	4	27	1	915
2005	759	34	4	12	1	810
2004	1079	55	0	28	1	1163
2003	1211	73	1	30	1	1316
2002	1305	52	5	31	0	1393
2001	868	34	2	26	0	930
2000	1391	50	8	27	0	1476
1999	1375	62	3	17	0	1457
1998	1299	35	1	19	4	1358
Total	14025	595	38	326	14	14998

Notes: *"?" means blank data field
1=Fair, 2=Rain, 3=Wind, 4=Fog
Source: ARI Accident Database 1998-2010

Table E-10. Year-wise double vehicle accidents vs light condition

Year	Light Condition					Total
	1	2	3	4	?*	
2010	499	95	52	65	0	711
2009	629	131	87	83	5	935
2008	885	185	105	110	1	1286
2007	878	155	109	98	8	1248
2006	667	94	82	69	3	915
2005	522	107	95	81	5	810
2004	768	152	101	138	4	1163
2003	864	201	98	150	3	1316
2002	962	156	162	113	0	1393
2001	619	135	85	89	2	930
2000	969	188	182	135	2	1476
1999	974	198	153	132	0	1457
1998	934	163	183	74	4	1358
Total	10170	1960	1494	1337	37	14998

Notes: *"?" means blank data field
1=Daylight, 2=Dawn/Dusk, 3=Night (lit), 4= Night (unlit)
Source: ARI Accident Database 1998-2010

Table E-11. Year-wise double vehicle accidents vs road geometric condition

Year	Road Geometric Condition						Total
	1	2	3	4	5	?*	
2010	634	47	14	16	0	0	711
2009	817	78	12	14	7	7	935
2008	1137	113	19	14	2	1	1286
2007	1122	86	10	12	4	14	1248
2006	839	61	6	5	0	4	915
2005	731	58	5	13	1	2	810
2004	1048	79	18	12	1	5	1163
2003	1197	92	8	10	2	7	1316
2002	1291	82	10	9	0	1	1393
2001	834	71	14	9	2	0	930
2000	1344	104	15	11	1	1	1476
1999	1345	89	11	8	3	1	1457
1998	1276	60	11	6	0	5	1358
Total	13615	1020	153	139	23	48	14998

Notes: *"?" means blank data field
1=Straight + Flat, 2=Curve only, 3=Slope only, 4=Curve + Slope, 5=Crest
Source: ARI Accident Database 1998-2010

Table E-12. Year-wise double vehicle accidents vs road surface condition

Year	Road Surface Condition						Total
	1	2	3	4	5	?*	
2010	688	18	0	0	5	0	711
2009	868	46	1	0	14	6	935
2008	1221	53	0	2	8	2	1286
2007	1153	66	4	0	10	15	1248
2006	884	27	0	0	1	3	915
2005	756	48	2	0	0	4	810
2004	1095	62	0	0	2	4	1163
2003	1229	76	0	2	6	3	1316
2002	1324	63	1	0	4	1	1393
2001	892	34	2	0	1	1	930
2000	1404	61	4	1	5	1	1476
1999	1395	58	0	0	4	0	1457
1998	1314	38	1	0	0	5	1358
Total	14223	650	15	5	60	45	14998

Notes: *"?" means blank data field
1=Dry, 2=Wet, 3=Muddy, 4=Flooded, 5=Other
Source: ARI Accident Database 1998-2010

Table E-13. Year-wise double vehicle accidents vs surface type

Year	Surface type				Total
	1	2	3	?*	
2010	706	2	3	0	711
2009	922	3	4	6	935
2008	1273	7	4	2	1286
2007	1231	5	1	11	1248
2006	907	3	0	5	915
2005	800	3	6	1	810
2004	1148	11	2	2	1163
2003	1303	7	2	4	1316
2002	1384	7	1	1	1393
2001	919	6	4	1	930
2000	1454	12	9	1	1476
1999	1443	9	5	0	1457
1998	1346	4	3	5	1358
Total	14836	79	44	39	14998

Notes: *"?" means blank data field
1=Sealed, 2=Brick, 3=Earth
Source: ARI Accident Database 1998-2010

Table E-14. Year-wise double vehicle accidents vs surface quality

Year	Surface Quality				Total
	1	2	3	?*	
2010	689	12	10	0	711
2009	884	35	10	6	935
2008	1241	32	9	4	1286
2007	1203	28	5	12	1248
2006	879	19	12	5	915
2005	782	15	8	5	810
2004	1115	17	29	2	1163
2003	1258	34	19	5	1316
2002	1363	23	6	1	1393
2001	893	26	9	2	930
2000	1440	18	18	0	1476
1999	1427	23	7	0	1457
1998	1339	12	2	5	1358
Total	14513	294	144	47	14998

Notes: *"?" means blank data field
1=Good, 2=Rough, 3=Under repair
Source: ARI Accident Database 1998-2010

Table E-15. Year-wise double vehicle accidents vs road class

Year	Road Class						Total
	1	2	3	4	5	?*	
2010	331	115	87	70	108	0	711
2009	447	142	134	69	142	1	935
2008	668	150	152	90	220	6	1286
2007	570	173	213	58	231	3	1248
2006	445	107	135	55	164	9	915
2005	314	140	65	61	229	1	810
2004	606	104	115	40	296	2	1163
2003	630	167	152	61	304	2	1316
2002	524	147	194	45	482	1	1393
2001	343	120	135	49	281	2	930
2000	595	154	173	47	506	1	1476
1999	591	126	158	54	525	3	1457
1998	397	103	101	31	718	8	1358
Total	6461	1748	1814	730	4206	39	14998

Notes: *"?" means blank data field
1=National, 2=Regional, 3=Feeder, 4=Rural road, 5=City
Source: ARI Accident Database 1998-2010

Table E-16. Year-wise double vehicle accidents vs road feature

Year	Road Feature						Total
	1	2	3	4	5	?*	
2010	678	21	5	6	0	1	711
2009	890	16	6	7	6	10	935
2008	1231	23	10	8	4	10	1286
2007	1177	23	2	10	4	32	1248
2006	870	26	2	3	1	13	915
2005	774	18	0	3	1	14	810
2004	1111	20	10	12	3	7	1163
2003	1267	23	8	9	4	5	1316
2002	1347	24	8	11	1	2	1393
2001	902	13	5	6	2	2	930
2000	1436	25	2	8	1	4	1476
1999	1421	15	7	9	4	1	1457
1998	1316	22	3	5	2	10	1358
Total	14420	269	68	97	33	111	14998

Notes: *"?" means blank data field
1=None, 2=Bridge, 3=Culvert, 4=Narrowing/Restriction, 5=Speed breakers
Source: ARI Accident Database 1998-2010

Table E-17. Year-wise double vehicle accidents vs road location

Year	Location			Total
	1	2	?*	
2010	270	431	10	711
2009	365	563	7	935
2008	509	759	18	1286
2007	506	712	30	1248
2006	322	565	28	915
2005	285	504	21	810
2004	399	750	14	1163
2003	447	855	14	1316
2002	571	816	6	1393
2001	368	559	3	930
2000	654	816	6	1476
1999	663	792	2	1457
1998	828	519	11	1358
Total	6187	8641	170	14998

Notes: *"?" means blank data field
1=Urban area, 2=Rural area
Source: ARI Accident Database 1998-2010

APPENDIX-F

Table F-1. Year-wise single vehicle accidents vs day of week

Year	Day of Week								?*
	1	2	3	4	5	6	7	Total	
2010	29	24	22	30	41	29	27	202	0
2009	47	36	42	42	55	30	32	284	0
2008	55	65	39	64	71	40	71	405	0
2007	59	71	60	74	66	61	75	466	0
2006	46	57	47	35	52	47	55	339	0
2005	49	52	45	58	40	51	40	335	0
2004	66	60	55	66	65	51	50	413	0
2003	83	79	87	72	84	90	78	573	0
2002	71	91	95	79	91	88	99	614	1
2001	45	64	61	65	70	70	60	435	0
2000	98	103	76	91	73	79	104	624	0
1999	86	85	102	88	68	107	82	618	0
1998	79	64	69	82	59	52	79	484	0
Total	813	851	800	846	835	795	852	5792	1

Notes: *"?" means blank data field
1=Monday, 2=Tuesday, 3=Wednesday, 4=Thursday, 5=Friday, 6=Saturday, 7=Sunday
Source: ARI Accident Database 1998-2010

Table F-2. Year-wise single vehicle accidents vs month of year

Year	Month of Year												Total
	1	2	3	4	5	6	7	8	9	10	11	12	
2010	22	18	16	19	18	5	19	15	23	11	24	12	202
2009	28	23	37	30	22	30	19	24	26	12	15	18	284
2008	44	30	45	27	56	33	40	35	18	29	17	31	405
2007	37	36	41	38	45	30	47	45	38	33	32	44	466
2006	30	35	30	40	35	29	33	21	20	24	21	21	339
2005	37	28	35	26	24	32	33	31	26	21	22	20	335
2004	42	42	37	40	34	35	32	25	35	30	29	32	413
2003	47	46	46	48	70	61	62	37	59	45	35	17	573
2002	52	48	66	57	79	52	49	37	40	29	47	59	615
2001	46	33	49	29	37	41	29	42	36	34	28	31	435
2000	61	61	50	56	62	50	62	52	39	51	39	41	624
1999	51	51	57	51	57	67	56	42	53	48	42	43	618
1998	46	51	51	46	41	51	45	33	25	30	27	38	484
Total	543	502	560	507	580	516	526	439	438	397	378	407	5793

Notes: 1=January, 2=February, 3=March, 4=April, 5=May, 6=June, 7=July, 8=August, 9=September, 10=October, 11=November, 12=December
Source: ARI Accident Database 1998-2010

Table F-3. Year-wise single vehicle accidents vs time of occurrence

Year	Time of Occurrence																									Total
	00-01	01-02	02-03	03-04	04-05	05-06	06-07	07-08	08-09	09-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17	17-18	18-19	19-20	20-21	21-22	22-23	23-24	?*	
2010	8	3	6	5	3	4	6	9	7	13	15	10	8	14	12	13	21	4	8	11	9	5	3	5	0	202
2009	10	5	6	11	17	11	11	12	9	18	14	21	14	16	9	17	16	21	7	10	10	7	3	8	1	284
2008	16	4	10	10	16	21	9	21	16	21	28	22	24	28	20	21	26	22	10	12	13	15	7	12	1	405
2007	7	7	8	22	12	12	17	21	22	27	29	41	28	18	28	32	32	26	12	12	14	13	5	15	6	466
2006	24	11	7	8	13	15	12	15	20	14	27	24	17	9	22	19	23	13	8	7	9	10	9	1	2	339
2005	45	9	8	9	11	13	19	24	18	20	12	19	21	15	12	18	15	10	6	9	5	6	4	4	3	335
2004	33	9	14	10	13	15	14	23	11	30	23	26	21	17	25	19	14	10	17	18	21	10	10	7	3	413
2003	10	20	12	16	20	28	25	18	21	42	51	41	31	23	37	29	26	21	18	20	16	17	13	9	9	573
2002	11	22	13	16	29	23	19	28	34	32	51	35	41	41	28	29	33	33	19	26	18	14	7	12	1	615
2001	4	12	13	14	11	17	15	15	20	20	29	34	20	24	19	31	27	26	10	19	14	16	13	12	0	435
2000	8	15	15	15	20	23	27	25	24	43	41	45	37	41	40	30	45	37	16	19	14	9	17	11	7	624
1999	7	13	10	14	21	24	14	21	35	41	41	41	50	41	35	35	34	33	24	21	17	6	13	24	3	618
1998	9	15	7	9	19	14	20	22	26	32	35	34	23	25	27	25	29	28	15	17	18	8	12	11	4	484
Total	192	145	129	159	205	220	208	254	263	353	396	393	335	312	314	318	341	284	170	201	178	136	116	131	40	5793

Notes: *"?" means blank data field
Source: ARI Accident Database 1998-2010

Table F-4. Year-wise single vehicle accidents vs junction type

Year	Junction Type								Total
	1	2	3	4	5	6	7	?*	
2010	147	13	7	5	0	0	30	0	202
2009	160	7	13	5	3	3	91	2	284
2008	262	27	22	5	1	0	86	2	405
2007	269	11	21	6	2	1	126	30	466
2006	207	8	18	6	3	1	80	16	339
2005	202	10	16	7	4	2	68	26	335
2004	294	9	26	1	2	2	68	11	413
2003	434	12	31	4	1	1	69	21	573
2002	524	15	25	4	3	1	40	3	615
2001	378	6	14	2	3	1	30	1	435
2000	550	10	27	10	1	0	25	1	624
1999	545	9	33	8	2	2	19	0	618
1998	381	15	36	3	6	0	31	12	484
Total	4353	152	289	66	31	14	763	125	5793

Notes: *"?" means blank data field

1=Not at junction, 2=Cross junction, 3=Tee junction, 4=Staggered tee junction, 5=Roundabouts, 6= Railway/level crossing, 7=Other

Source: ARI Accident Database 1998-2010

Table F-5. Year-wise single vehicle accidents vs traffic control system

Year	Traffic Control									Total
	1	2	3	4	5	6	7	8	?*	
2010	161	2	0	9	4	0	0	26	0	202
2009	232	2	1	4	3	2	2	34	4	284
2008	314	4	2	11	3	4	2	58	7	405
2007	327	5	0	18	2	1	2	86	25	466
2006	259	5	2	13	6	2	1	40	11	339
2005	252	7	2	13	6	5	3	36	11	335
2004	336	2	1	8	8	7	3	39	9	413
2003	492	10	3	17	3	2	0	38	8	573
2002	536	8	3	19	6	4	1	36	2	615
2001	387	6	1	8	5	4	4	19	1	435
2000	562	10	1	16	4	6	1	24	0	624
1999	554	20	0	14	1	12	4	13	0	618
1998	382	19	0	11	2	25	2	27	16	484
Total	4794	100	16	161	53	74	25	476	94	5793

Notes: *"?" means blank data field
1=No control, 2=Centerline marking, 3=Pedestrian crossing, 4=Police controlled, 5=Traffic lights, 6=Police + Traffic lights, 7=Stop/Give way sign, 8=Other
Source: ARI Accident Database 1998-2010

Table F-6. Year-wise single vehicle accidents vs traffic collision type

Year	Collision Type												Total
	1	2	3	4	5	6	7	8	9	10	11	?*	
2010	4	7	0	4	125	8	24	0	0	0	30	0	202
2009	8	8	1	5	166	10	36	5	0	0	45	0	284
2008	15	19	2	12	220	15	56	1	0	0	64	1	405
2007	8	23	2	21	242	24	45	6	0	3	90	2	466
2006	12	16	2	13	164	20	47	4	0	1	57	3	339
2005	5	12	2	14	159	22	54	4	0	2	58	3	335
2004	15	12	0	28	219	9	49	2	0	1	74	4	413
2003	8	8	0	12	342	13	84	5	0	3	94	4	573
2002	5	8	0	10	372	22	87	11	0	1	99	0	615
2001	7	7	0	5	253	13	56	3	0	1	90	0	435
2000	2	3	0	6	328	21	108	3	0	2	151	0	624
1999	6	8	0	12	334	26	112	1	0	1	118	0	618
1998	7	13	0	9	257	23	64	3	0	3	100	5	484
Total	102	144	9	151	3181	226	822	48	0	18	1070	22	5793

Notes: *"?" means blank data field
1=Head on, 2=Rear end, 3=Right angle, 4=Side swipe, 5=Overturn, 6=Hit object in road, 7=Hit object off road, 8=Hit parked vehicle, 9=Hit pedestrian, 10=Hit animal, 11=Other
Source: ARI Accident Database 1998-2010

Table F-7. Year-wise single vehicle accidents vs traffic movement

Year	Traffic Movement			Total
	1	2	?*	
2010	16	186	0	202
2009	30	250	4	284
2008	60	342	3	405
2007	129	322	15	466
2006	60	275	4	339
2005	44	283	8	335
2004	56	350	7	413
2003	68	500	5	573
2002	36	579	0	615
2001	18	417	0	435
2000	31	593	0	624
1999	22	595	1	618
1998	31	452	1	484
Total	601	5144	48	5793

Notes: *"?" means blank data field
1=1-Way street, 2=2-Way street
Source: ARI Accident Database 1998-2010

Table F-8. Year-wise single vehicle accidents vs presence of divider in roads

Year	Presence of Divider			Total
	1	2	?*	
2010	24	178	0	202
2009	20	259	5	284
2008	46	351	8	405
2007	56	380	30	466
2006	43	282	14	339
2005	38	279	18	335
2004	47	343	23	413
2003	44	508	21	573
2002	67	548	0	615
2001	45	390	0	435
2000	73	550	1	624
1999	59	557	2	618
1998	81	398	5	484
Total	643	5023	127	5793

Notes: *"?" means blank data field
1=Yes, 2=No
Source: ARI Accident Database 1998-2010

Table F-9. Year-wise single vehicle accidents vs weather condition

Year	Weather Condition					Total
	1	2	3	4	?*	
2010	191	5	0	6	0	202
2009	258	14	0	12	0	284
2008	361	25	1	18	0	405
2007	416	33	2	9	6	466
2006	309	12	2	13	3	339
2005	302	20	1	9	3	335
2004	374	26	0	13	0	413
2003	500	50	4	18	1	573
2002	541	53	1	20	0	615
2001	395	30	0	10	0	435
2000	569	40	2	13	0	624
1999	557	46	3	12	0	618
1998	440	29	0	14	1	484
Total	5213	383	16	167	14	5793

Notes: *“?” means blank data field
1=Fair, 2=Rain, 3=Wind, 4=Fog
Source: ARI Accident Database 1998-2010

Table F-10. Year-wise single vehicle accidents vs light condition

Year	Light Condition					Total
	1	2	3	4	?*	
2010	132	30	7	33	0	202
2009	180	42	19	41	2	284
2008	288	55	18	43	1	405
2007	319	49	34	54	10	466
2006	224	44	21	45	5	339
2005	201	72	16	43	3	335
2004	237	77	31	66	2	413
2003	351	104	31	84	3	573
2002	388	96	37	94	0	615
2001	258	70	20	87	0	435
2000	413	95	38	78	0	624
1999	410	83	31	94	0	618
1998	316	61	35	69	3	484
Total	3717	878	338	831	29	5793

Notes: *“?” means blank data field
1=Daylight, 2=Dawn/Dusk, 3=Night (lit), 4= Night (unlit)
Source: ARI Accident Database 1998-2010

Table F-11. Year-wise single vehicle accidents vs road geometric condition

Year	Road Geometric Condition						Total
	1	2	3	4	5	?*	
2010	168	22	3	7	2	0	202
2009	229	25	6	18	4	2	284
2008	307	46	21	25	5	1	405
2007	381	42	15	12	5	11	466
2006	276	40	8	8	2	5	339
2005	257	34	11	13	14	6	335
2004	355	27	9	13	7	2	413
2003	494	49	6	16	5	3	573
2002	534	53	13	5	8	2	615
2001	356	48	18	9	4	0	435
2000	521	57	20	19	7	0	624
1999	508	57	27	17	9	0	618
1998	393	50	17	14	7	3	484
Total	4779	550	174	176	79	35	5793

Notes: *"?" means blank data field
1=Straight + Flat, 2=Curve only, 3=Slope only, 4=Curve + Slope, 5=Crest
Source: ARI Accident Database 1998-2010

Table F-12. Year-wise single vehicle accidents vs road surface condition

Year	Road Surface Condition						Total
	1	2	3	4	5	?*	
2010	185	8	0	0	9	0	202
2009	254	18	1	1	7	3	284
2008	358	36	3	0	8	0	405
2007	418	31	3	1	5	8	466
2006	311	17	3	1	2	5	339
2005	304	15	5	1	3	7	335
2004	382	22	3	0	5	1	413
2003	510	53	5	0	3	2	573
2002	545	62	2	0	4	2	615
2001	400	31	1	0	3	0	435
2000	575	40	3	0	6	0	624
1999	560	51	3	1	3	0	618
1998	440	35	4	0	3	2	484
Total	5242	419	36	5	61	30	5793

Notes: *"?" means blank data field
1=Dry, 2=Wet, 3=Muddy, 4=Flooded, 5=Other
Source: ARI Accident Database 1998-2010

Table F-13. Year-wise single vehicle accidents vs surface type

Year	Surface type				Total
	1	2	3	?*	
2010	194	4	4	0	202
2009	265	7	10	2	284
2008	388	8	9	0	405
2007	439	13	5	9	466
2006	314	9	11	5	339
2005	316	8	7	4	335
2004	390	9	13	1	413
2003	545	11	13	4	573
2002	587	15	11	2	615
2001	418	7	10	0	435
2000	593	18	12	1	624
1999	595	11	12	0	618
1998	460	12	11	1	484
Total	5504	132	128	29	5793

Notes: *“?” means blank data field
1=Sealed, 2=Brick, 3=Earth
Source: ARI Accident Database 1998-2010

Table F-14. Year-wise single vehicle accidents vs surface quality

Year	Surface Quality				Total
	1	2	3	?*	
2010	183	16	3	0	202
2009	260	14	8	2	284
2008	357	37	9	2	405
2007	417	27	11	11	466
2006	307	19	7	6	339
2005	286	36	8	5	335
2004	360	22	29	2	413
2003	504	42	23	4	573
2002	544	49	20	2	615
2001	390	34	10	1	435
2000	569	36	19	0	624
1999	571	38	9	0	618
1998	449	28	6	1	484
Total	5197	398	162	36	5793

Notes: *“?” means blank data field
1=Good, 2=Rough, 3=Under repair
Source: ARI Accident Database 1998-2010

Table F-15. Year-wise single vehicle accidents vs road class

Year	Road Class						Total
	1	2	3	4	5	?*	
2010	73	41	34	31	23	0	202
2009	123	39	54	47	21	0	284
2008	180	58	77	66	24	0	405
2007	209	54	109	52	42	0	466
2006	150	39	66	42	39	3	339
2005	153	48	55	50	28	1	335
2004	207	64	67	36	36	3	413
2003	268	97	108	52	43	5	573
2002	250	86	152	56	71	0	615
2001	180	62	94	49	49	1	435
2000	251	76	151	57	88	1	624
1999	274	104	128	40	71	1	618
1998	165	64	93	52	105	5	484
Total	2483	832	1188	630	640	20	5793

Notes: *"?" means blank data field
1=National, 2=Regional, 3=Feeder, 4=Rural road, 5=City
Source: ARI Accident Database 1998-2010

Table F-16. Year-wise single vehicle accidents vs road feature

Year	Road Feature						Total
	1	2	3	4	5	?*	
2010	192	7	1	1	1	0	202
2009	263	9	2	5	3	2	284
2008	368	14	6	12	3	2	405
2007	427	10	7	4	0	18	466
2006	309	12	5	5	0	8	339
2005	293	10	7	10	0	15	335
2004	382	12	5	10	1	3	413
2003	533	15	7	12	3	3	573
2002	574	12	10	16	2	1	615
2001	400	15	9	8	3	0	435
2000	578	22	9	12	3	0	624
1999	578	20	8	8	2	2	618
1998	451	11	4	12	0	6	484
Total	5348	169	80	115	21	60	5793

Notes: *"?" means blank data field
1=None, 2=Bridge, 3=Culvert, 4=Narrowing/Restriction, 5=Speed breakers
Source: ARI Accident Database 1998-2010

Table F-17. Year-wise single vehicle accidents vs road location

Year	Location			Total
	1	2	?*	
2010	54	147	1	202
2009	60	221	3	284
2008	99	301	5	405
2007	134	313	19	466
2006	82	240	17	339
2005	46	274	15	335
2004	66	341	6	413
2003	85	483	5	573
2002	102	511	2	615
2001	70	360	5	435
2000	136	486	2	624
1999	103	512	3	618
1998	128	349	7	484
Total	1165	4538	90	5793

Notes: *"?" means blank data field
1=Urban area, 2=Rural area
Source: ARI Accident Database 1998-2010

APPENDIX-G

Table G-1. Year-wise multi vehicle accidents vs day of week

Year	Day of Week								?*
	1	2	3	4	5	6	7	Total	
2010	1	0	0	3	6	2	5	17	0
2009	3	2	1	4	3	3	3	19	0
2008	2	0	2	2	3	1	0	10	0
2007	3	1	1	1	1	1	2	10	0
2006	1	1	1	1	1	0	0	5	0
2005	1	2	2	0	1	2	0	8	0
2004	3	3	4	4	4	4	3	25	0
2003	2	1	1	2	3	1	4	14	0
2002	0	1	2	1	0	0	2	6	0
2001	1	3	1	1	2	0	1	9	0
2000	2	3	4	3	1	3	2	18	0
1999	5	2	2	6	4	5	3	27	0
1998	5	5	7	3	4	5	9	38	0
Total	29	24	28	31	33	27	34	206	0

Notes: *"?" means blank data field
1=Monday, 2=Tuesday, 3=Wednesday, 4=Thursday, 5=Friday, 6=Saturday, 7=Sunday
Source: ARI Accident Database 1998-2010

Table G-2. Year-wise multi vehicle accidents vs month of year

Year	Month of Year												Total
	1	2	3	4	5	6	7	8	9	10	11	12	
2010	0	2	1	3	2	3	1	1	2	2	0	0	17
2009	1	1	1	1	0	2	3	3	3	2	0	2	19
2008	1	0	2	1	0	0	2	2	0	1	1	0	10
2007	0	0	3	3	0	1	0	0	0	1	1	1	10
2006	0	2	0	1	1	0	0	0	0	0	0	1	5
2005	1	1	1	1	2	0	1	1	0	0	0	0	8
2004	4	2	0	3	1	5	0	1	0	2	2	5	25
2003	2	1	1	1	1	1	1	2	0	2	2	0	14
2002	0	0	1	1	0	0	0	1	0	1	2	0	6
2001	0	0	1	0	0	0	0	1	2	2	1	2	9
2000	2	0	7	2	0	1	0	1	3	0	1	1	18
1999	1	4	2	3	1	1	1	4	0	3	5	2	27
1998	3	2	1	2	4	6	2	5	2	4	4	3	38
Total	15	15	21	22	12	20	11	22	12	20	19	17	206

Notes: 1=January, 2=February, 3=March, 4=April, 5=May, 6=June, 7=July, 8=August, 9=September, 10=October, 11=November, 12=December
Source: ARI Accident Database 1998-2010

Table G-3. Year-wise multi vehicle accidents vs time of occurrence

Year	Time of Occurrence																								Total
	00-01	01-02	02-03	03-04	04-05	05-06	06-07	07-08	08-09	09-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17	17-18	18-19	19-20	20-21	21-22	22-23	23-24	?*	
2010	0	0	0	0	2	0	0	0	1	0	1	1	1	2	1	1	1	1	1	1	1	1	0	1	0	17
2009	0	1	1	0	0	1	0	1	0	0	1	1	3	3	0	1	2	0	1	0	0	0	0	3	0	19
2008	0	0	0	0	0	1	0	0	0	1	3	0	0	1	0	0	1	2	0	0	0	1	0	0	0	10
2007	2	0	1	1	0	0	0	0	0	1	0	0	2	0	1	1	0	1	0	0	0	0	0	0	0	10
2006	1	0	0	0	0	1	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5
2005	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	2	2	1	0	0	0	0	0	0	8
2004	3	0	0	1	0	0	1	0	3	3	3	0	6	1	1	0	2	0	1	0	0	0	0	0	0	25
2003	0	0	0	1	0	0	1	0	2	1	1	1	2	0	3	1	0	0	0	0	0	0	0	1	0	14
2002	0	0	0	0	0	0	0	0	1	0	1	1	0	1	0	0	0	0	0	2	0	0	0	0	0	6
2001	0	0	0	0	0	0	0	1	1	0	1	0	3	1	0	0	0	0	1	0	0	1	0	0	0	9
2000	0	1	0	1	1	0	1	1	0	0	2	2	2	1	1	0	0	1	1	0	1	0	1	1	0	18
1999	0	0	1	0	0	2	0	1	1	2	1	0	2	0	2	2	3	0	3	3	1	2	0	1	0	27
1998	0	0	0	1	0	0	1	1	1	4	4	5	4	5	3	0	2	1	1	1	1	1	2	0	0	38
Total	6	3	3	5	3	5	5	7	10	12	19	11	25	15	12	7	13	8	10	7	4	6	3	7	0	206

Notes: *"?" means blank data field
Source: ARI Accident Database 1998-2010

Table G-4. Year-wise multi vehicle accidents vs junction type

Year	Junction Type								Total
	1	2	3	4	5	6	7	?*	
2010	13	3	1	0	0	0	0	0	17
2009	7	4	5	0	0	0	3	0	19
2008	7	1	0	0	1	0	1	0	10
2007	9	0	0	0	0	0	1	0	10
2006	3	1	1	0	0	0	0	0	5
2005	5	0	0	0	0	2	0	1	8
2004	16	2	4	0	1	0	1	1	25
2003	6	1	4	0	0	0	2	1	14
2002	4	0	1	0	0	1	0	0	6
2001	8	0	1	0	0	0	0	0	9
2000	10	2	6	0	0	0	0	0	18
1999	17	4	5	0	1	0	0	0	27
1998	24	3	6	0	3	2	0	0	38
Total	129	21	34	0	6	5	8	3	206

Notes: *"?" means blank data field
1=Not at junction, 2=Cross junction, 3=Tee junction, 4=Staggered tee junction, 5=Roundabouts, 6= Railway/level crossing, 7=Other
Source: ARI Accident Database 1998-2010

Table G-5. Year-wise multi vehicle accidents vs traffic control system

Year	Traffic Control									Total
	1	2	3	4	5	6	7	8	?*	
2010	14	0	0	3	0	0	0	0	0	17
2009	10	0	0	5	2	1	0	1	0	19
2008	7	0	0	1	0	0	1	0	1	10
2007	10	0	0	0	0	0	0	0	0	10
2006	5	0	0	0	0	0	0	0	0	5
2005	6	0	0	1	0	1	0	0	0	8
2004	20	1	0	2	0	1	0	1	0	25
2003	7	1	0	3	0	3	0	0	0	14
2002	5	0	0	0	0	0	0	1	0	6
2001	7	1	0	0	0	1	0	0	0	9
2000	14	1	0	0	0	3	0	0	0	18
1999	19	0	0	0	0	7	0	1	0	27
1998	24	4	0	5	0	5	0	0	0	38
Total	148	8	0	20	2	22	1	4	1	206

Notes: *"?" means blank data field

1=No control, 2=Centerline marking, 3=Pedestrian crossing, 4=Police controlled, 5=Traffic lights, 6=Police + Traffic lights, 7=Stop/Give way sign, 8=Other

Source: ARI Accident Database 1998-2010

Table G-6. Year-wise multi vehicle accidents vs traffic collision type

Year	Collision Type												Total
	1	2	3	4	5	6	7	8	9	10	11	?*	
2010	8	7	0	1	0	0	0	1	0	0	0	0	17
2009	6	8	0	1	0	1	0	1	0	1	1	0	19
2008	2	6	0	1	0	0	0	1	0	0	0	0	10
2007	3	4	0	2	1	0	0	0	0	0	0	0	10
2006	1	3	0	1	0	0	0	0	0	0	0	0	5
2005	1	7	0	0	0	0	0	0	0	0	0	0	8
2004	10	8	1	1	0	0	0	4	0	0	1	0	25
2003	4	6	0	0	1	0	0	3	0	0	0	0	14
2002	1	5	0	0	0	0	0	0	0	0	0	0	6
2001	3	6	0	0	0	0	0	0	0	0	0	0	9
2000	5	7	2	3	0	0	0	1	0	0	0	0	18
1999	2	20	0	4	0	0	0	1	0	0	0	0	27
1998	7	18	1	8	0	1	0	3	0	0	0	0	38
Total	53	105	4	22	2	2	0	15	0	1	2	0	206

Notes: *"?" means blank data field
1=Head on, 2=Rear end, 3=Right angle, 4=Side swipe, 5=Overturn, 6=Hit object in road, 7=Hit object off road, 8=Hit parked vehicle, 9=Hit pedestrian, 10=Hit animal, 11=Other
Source: ARI Accident Database 1998-2010

Table G-7. Year-wise multi vehicle accidents vs traffic movement

Year	Traffic Movement			Total
	1	2	?*	
2010	1	16	0	17
2009	7	12	0	19
2008	4	6	0	10
2007	1	9	0	10
2006	1	4	0	5
2005	0	8	0	8
2004	1	24	0	25
2003	1	13	0	14
2002	1	5	0	6
2001	0	9	0	9
2000	1	17	0	18
1999	0	27	0	27
1998	3	35	0	38
Total	21	185	0	206

Notes: *"?" means blank data field
1=1-Way street, 2=2-Way street
Source: ARI Accident Database 1998-2010

Table G-8. Year-wise multi vehicle accidents vs presence of divider in roads

Year	Presence of Divider			Total
	1	2	?*	
2010	6	11	0	17
2009	9	10	0	19
2008	4	6	0	10
2007	3	7	0	10
2006	2	3	0	5
2005	5	3	0	8
2004	6	16	3	25
2003	8	5	1	14
2002	4	2	0	6
2001	5	4	0	9
2000	7	11	0	18
1999	16	11	0	27
1998	22	16	0	38
Total	97	105	4	206

Notes: *"?" means blank data field
1=Yes, 2=No
Source: ARI Accident Database 1998-2010

Table G-9. Year-wise multi vehicle accidents vs weather condition

Year	Weather Condition					Total
	1	2	3	4	?*	
2010	17	0	0	0	0	17
2009	17	1	0	1	0	19
2008	10	0	0	0	0	10
2007	8	1	0	1	0	10
2006	3	1	0	1	0	5
2005	8	0	0	0	0	8
2004	24	1	0	0	0	25
2003	13	0	0	1	0	14
2002	6	0	0	0	0	6
2001	9	0	0	0	0	9
2000	16	1	0	1	0	18
1999	25	2	0	0	0	27
1998	37	0	0	1	0	38
Total	193	7	0	6	0	206

Notes: *"?" means blank data field
1=Fair, 2=Rain, 3=Wind, 4=Fog
Source: ARI Accident Database 1998-2010

Table G-10. Year-wise multi vehicle accidents vs light condition

Year	Light Condition					Total
	1	2	3	4	?*	
2010	11	2	3	1	0	17
2009	11	2	4	2	0	19
2008	7	1	1	1	0	10
2007	8	0	1	1	0	10
2006	2	3	0	0	0	5
2005	7	1	0	0	0	8
2004	19	4	0	2	0	25
2003	11	1	1	1	0	14
2002	5	0	1	0	0	6
2001	7	1	1	0	0	9
2000	10	3	3	2	0	18
1999	15	7	5	0	0	27
1998	31	2	5	0	0	38
Total	144	27	25	10	0	206

Notes: *"?" means blank data field
1=Daylight, 2=Dawn/Dusk, 3=Night (lit), 4= Night (unlit)
Source: ARI Accident Database 1998-2010

Table G-11. Year-wise multi vehicle accidents vs road geometric condition

Year	Road Geometric Condition						Total
	1	2	3	4	5	?*	
2010	14	2	1	0	0	0	17
2009	15	4	0	0	0	0	19
2008	7	1	1	1	0	0	10
2007	10	0	0	0	0	0	10
2006	5	0	0	0	0	0	5
2005	7	1	0	0	0	0	8
2004	24	0	1	0	0	0	25
2003	13	1	0	0	0	0	14
2002	5	1	0	0	0	0	6
2001	9	0	0	0	0	0	9
2000	16	2	0	0	0	0	18
1999	27	0	0	0	0	0	27
1998	37	1	0	0	0	0	38
Total	189	13	3	1	0	0	206

Notes: *"?" means blank data field
1=Straight + Flat, 2=Curve only, 3=Slope only, 4=Curve + Slope, 5=Crest
Source: ARI Accident Database 1998-2010

Table G-12. Year-wise multi vehicle accidents vs road surface condition

Year	Road Surface Condition						Total
	1	2	3	4	5	?*	
2010	17	0	0	0	0	0	17
2009	18	1	0	0	0	0	19
2008	10	0	0	0	0	0	10
2007	9	1	0	0	0	0	10
2006	5	0	0	0	0	0	5
2005	8	0	0	0	0	0	8
2004	24	1	0	0	0	0	25
2003	14	0	0	0	0	0	14
2002	6	0	0	0	0	0	6
2001	9	0	0	0	0	0	9
2000	17	1	0	0	0	0	18
1999	25	2	0	0	0	0	27
1998	38	0	0	0	0	0	38
Total	200	6	0	0	0	0	206

Notes: *"?" means blank data field
1=Dry, 2=Wet, 3=Muddy, 4=Flooded, 5=Other
Source: ARI Accident Database 1998-2010

Table G-13. Year-wise multi vehicle accidents vs surface type

Year	Surface type				Total
	1	2	3	?*	
2010	17	0	0	0	17
2009	19	0	0	0	19
2008	10	0	0	0	10
2007	10	0	0	0	10
2006	5	0	0	0	5
2005	8	0	0	0	8
2004	25	0	0	0	25
2003	14	0	0	0	14
2002	6	0	0	0	6
2001	9	0	0	0	9
2000	18	0	0	0	18
1999	27	0	0	0	27
1998	38	0	0	0	38
Total	206	0	0	0	206

Notes: *"?" means blank data field
1=Sealed, 2=Brick, 3=Earth
Source: ARI Accident Database 1998-2010

Table G-14. Year-wise multi vehicle accidents vs surface quality

Year	Surface Quality				Total
	1	2	3	?*	
2010	17	0	0	0	17
2009	19	0	0	0	19
2008	10	0	0	0	10
2007	10	0	0	0	10
2006	5	0	0	0	5
2005	8	0	0	0	8
2004	25	0	0	0	25
2003	14	0	0	0	14
2002	6	0	0	0	6
2001	9	0	0	0	9
2000	18	0	0	0	18
1999	27	0	0	0	27
1998	38	0	0	0	38
Total	206	0	0	0	206

Notes: *"?" means blank data field
1=Good, 2=Rough, 3=Under repair
Source: ARI Accident Database 1998-2010

Table G-15. Year-wise multi vehicle accidents vs road class

Year	Road Class						Total
	1	2	3	4	5	?*	
2010	11	2	0	0	4	0	17
2009	12	0	1	0	6	0	19
2008	5	1	0	0	4	0	10
2007	7	0	1	0	2	0	10
2006	3	0	2	0	0	0	5
2005	2	1	0	0	5	0	8
2004	20	1	0	0	4	0	25
2003	4	1	0	0	9	0	14
2002	1	0	0	0	5	0	6
2001	2	0	0	0	7	0	9
2000	6	0	0	0	12	0	18
1999	0	2	0	0	25	0	27
1998	6	0	1	0	31	0	38
Total	79	8	5	0	114	0	206

Notes: *"?" means blank data field
1=National, 2=Regional, 3=Feeder, 4=Rural road, 5=City
Source: ARI Accident Database 1998-2010

Table G-16. Year-wise multi vehicle accidents vs road feature

Year	Road Feature						Total
	1	2	3	4	5	?*	
2010	16	0	0	0	1	0	17
2009	18	1	0	0	0	0	19
2008	9	0	0	1	0	0	10
2007	10	0	0	0	0	0	10
2006	4	0	0	0	0	1	5
2005	8	0	0	0	0	0	8
2004	23	1	1	0	0	0	25
2003	14	0	0	0	0	0	14
2002	6	0	0	0	0	0	6
2001	9	0	0	0	0	0	9
2000	18	0	0	0	0	0	18
1999	27	0	0	0	0	0	27
1998	38	0	0	0	0	0	38
Total	200	2	1	1	1	1	206

Notes: *"?" means blank data field
1=None, 2=Bridge, 3=Culvert, 4=Narrowing/Restriction, 5=Speed breakers
Source: ARI Accident Database 1998-2010

Table G-17. Year-wise multi vehicle accidents vs road location

Year	Location			Total
	1	2	?*	
2010	11	5	1	17
2009	12	6	1	19
2008	6	4	0	10
2007	3	7	0	10
2006	1	3	1	5
2005	5	3	0	8
2004	8	17	0	25
2003	10	4	0	14
2002	5	1	0	6
2001	7	2	0	9
2000	13	5	0	18
1999	26	1	0	27
1998	33	4	1	38
Total	140	62	4	206

Notes: *"?" means blank data field
1=Urban area, 2=Rural area
Source: ARI Accident Database 1998-2010

APPENDIX-H

Table H-1. Summary of variable importance for pedestrian accidents cluster 1

Variable	Mean Decrease Accuracy
RdClassM	26.9062199
LightM	26.83743161
CollTypeM	23.72993413
LocatTypeM	22.5120291
Time_SQL	16.75631434
DividerM	16.58498867
MovM	13.00608565
No_DrvCasu	12.6085845
TrafficContrlM	12.25567782
JuncTypeM	8.86385692
No_PedCasu	7.99961349
Rd_GeoM	6.871664663
No_VehInv	6.319291629
RdFeatuM	4.848950542
Month	4.674958225
Surf_QualM	4.659464214
No_PassCasu	4.64026838
Surf_TypeM	3.141168447
WeatherM	3.043011884
Surf_CondM	1.674016489
Day	0.940054443

Table H-2. Summary of variable importance for pedestrian accidents cluster 2

Variable	Mean Decrease Accuracy
LocatTypeM	8.934941283
No_PedCasu	8.269677705
TrafficContrlM	5.746201222
RdClassM	4.269378045
JuncTypeM	4.148095324
Month	2.385018256
DividerM	2.225364847
WeatherM	1.895197068
Time_SQL	1.8325482
Rd_GeoM	1.46896939
Surf_TypeM	1.356936071
Day	0.705877594
No_VehInv	0
No_DrvCasu	0
No_PassCasu	0
RdFeatuM	-0.305323702
Surf_QualM	-0.336914902
MovM	-0.491608575
CollTypeM	-0.851432457
Surf_CondM	-1.009755478
LightM	-1.305656273

Table H-3. Summary of variable importance for pedestrian accidents cluster 3

Variable	Mean Decrease Accuracy
TrafficContrlM	4.783570879
Time_SQL	2.558916581
Day	2.483832145
JuncTypeM	1.733000617
Month	1.720957291
RdClassM	1.268245094
Surf_TypeM	1.155791833
WeatherM	1.055525191
No_DrvCasu	1.001001503
LightM	0.968222877
Surf_CondM	0.735977243
MovM	0.527484146
Surf_QualM	0.364850445
No_PedCasu	0.096436748
RdFeatuM	0.001048796
No_VehInv	0
No_PassCasu	0
DividerM	-0.034176147
LocatTypeM	-1.245013484
CollTypeM	-1.739744352
Rd_GeoM	-2.06444957

Table H-4. Summary of variable importance for pedestrian accidents cluster 4

Variable	Mean Decrease Accuracy
JuncTypeM	9.237595977
Month	5.463137275
Time_SQL	5.133861743
Day	4.449203544
No_DrvCasu	3.766832455
RdClassM	3.237877133
TrafficContrlM	3.188050255
CollTypeM	2.881890016
No_VehInv	2.562102166
No_PassCasu	2.332402853
DividerM	1.913509419
LocatTypeM	1.908486738
Surf_CondM	1.75696986
Rd_GeoM	1.754408227
LightM	1.477500111
No_PedCasu	0.998180088
RdFeatuM	-0.322919388
WeatherM	-0.390519321
Surf_QualM	-0.507383819
MovM	-2.435402268
Surf_TypeM	-4.32681388

Table H-5. Summary of variable importance for double vehicle accidents cluster 1

Variable	Mean Decrease Accuracy
No_PassCasu	22.64081763
No_DrvCasu	22.12735771
Rd_GeoM	10.30768083
LightM	6.152270616
CollTypeM	6.039192418
Surf_TypeM	5.525508339
TrafficContrlM	4.80883211
Time_SQL	3.902070387
RdClassM	3.805565585
Surf_CondM	3.136800618
Month	2.933717775
Day	2.930634923
JuncTypeM	2.199479313
LocatTypeM	1.454540657
Surf_QualM	1.137793456
DividerM	0.094012485
No_VehInv	0
No_PedCasu	0
MovM	-0.903885376
WeatherM	-1.013491735
RdFeatuM	-2.593078695

Table H-6. Summary of variable importance for double vehicle accidents cluster 4

Variable	Mean Decrease Accuracy
No_DrvCasu	5.851814411
No_PassCasu	4.344697098
RdClassM	3.184766888
CollTypeM	3.148147297
Surf_CondM	2.011488866
LocatTypeM	1.615001565
Rd_GeoM	1.532404265
LightM	1.119473278
Surf_QualM	0.913590845
DividerM	0.430048576
JuncTypeM	0.410031189
No_VehInv	0
No_PedCasu	0
MovM	-0.34708825
Surf_TypeM	-0.369043765
WeatherM	-0.491501564
Month	-0.567723234
Day	-0.813846022
TrafficContrlM	-1.010855469
RdFeatuM	-1.030854857
Time_SQL	-1.472380002

Table H-7. Summary of variable importance for single vehicle accidents cluster 1

Variable	Mean Decrease Accuracy
No_PassCasu	43.41367245
No_DrvCasu	21.42364882
CollTypeM	8.697937388
Time_SQL	6.70632163
LightM	6.407632429
Rd_GeoM	6.195833895
TrafficContrlM	4.119314096
JuncTypeM	3.779245973
WeatherM	2.197300965
Day	2.002030734
Month	1.50351611
Surf_TypeM	1.232597633
LocatTypeM	0.536435165
Surf_CondM	0.223043916
Surf_QualM	0.203889904
No_VehInv	0
No_PedCasu	0
RdClassM	-0.535777258
RdFeatuM	-1.494540202
DividerM	-1.662103406
MovM	-2.639128796

Table H-8. Summary of variable importance for single vehicle accidents cluster 3

Variable	Mean Decrease Accuracy
Day	4.790790311
Surf_TypeM	4.09891818
LightM	3.497018816
JuncTypeM	3.413011471
No_PassCasu	2.836218648
WeatherM	2.282952542
RdFeatuM	1.915971568
Surf_CondM	1.854059589
Surf_QualM	1.457035957
Month	1.388462211
Rd_GeoM	1.165519654
RdClassM	1.121842407
Time_SQL	0.879376819
TrafficContrlM	0.443102569
LocatTypeM	0.351178253
No_VehInv	0
No_PedCasu	0
MovM	-0.008413672
CollTypeM	-0.719038396
No_DrvCasu	-1.154396285
DividerM	-1.165600468

Table H-9. Summary of variable importance for single vehicle accidents cluster 4

Variable	Mean Decrease Accuracy
No_PassCasu	6.735444514
CollTypeM	5.53668368
MovM	4.982673333
No_DrvCasu	4.056968574
Surf_CondM	3.452905007
Month	3.279688221
Day	2.678137594
JuncTypeM	2.391014548
DividerM	2.318354462
WeatherM	1.667910494
Surf_QualM	1.15267288
Time_SQL	1.086494071
LocatTypeM	0.394505082
No_VehInv	0
No_PedCasu	0
LightM	-0.210289317
TrafficContrlM	-0.424045592
Rd_GeoM	-0.506852453
RdClassM	-0.689719124
RdFeatuM	-0.799988131
Surf_TypeM	-1.33869404

Table H-10. Summary of variable importance for multi vehicle accidents cluster 2

Variable	Mean Decrease Accuracy
No_PassCasu	3.5198123
Month	1.638168072
LocatTypeM	1.051854514
Surf_CondM	0.926921057
Rd_GeoM	0.229327544
No_VehInv	0
No_PedCasu	0
TrafficContrlM	0
MovM	0
DividerM	0
Surf_TypeM	0
Surf_QualM	0
RdFeatuM	0
RdClassM	-0.071104906
LightM	-0.0851073
JuncTypeM	-0.275292227
CollTypeM	-1.144290596
Time_SQL	-1.644682903
No_DrvCasu	-1.668027182
WeatherM	-1.9806105
Day	-2.5072853

Table H-11. Summary of variable importance for multi vehicle accidents cluster 3

Variable	Mean Decrease Accuracy
No_DrvCasu	7.047745792
No_PassCasu	6.298708301
Time_SQL	3.207650584
LightM	2.992148949
Rd_GeoM	1.784178541
Day	1.669843238
RdClassM	0.648067197
No_PedCasu	0
TrafficContrlM	0
MovM	0
WeatherM	0
Surf_CondM	0
Surf_TypeM	0
Surf_QualM	0
RdFeatuM	0
No_VehInv	-0.335710372
DividerM	-1.002760136
CollTypeM	-1.024357774
Month	-1.4416816
LocatTypeM	-2.172437226
JuncTypeM	-3.304094153