INJURY SEVERITY AMONG PEDESTRIANS IN METROPOLITAN CITIES : A DATA MINING APPROACH

SUBMITTED TO IITM

A THESIS

submitted by

KARTHIKEYAN R

for the award of the degree

of

Master of Technology

&

Bachelor of Technology

DEPARTMENT OF ENGINEERING DESIGN

INDIAN INSTITUTE OF TECHNOLOGY MADRAS.

MAY 2018

THESIS CERTIFICATE

This is to certify that the thesis titled Injury Severity among Pedestrians in Metropolitan Cities: A Data Mining Approach, submitted by **Karthikeyan R[ED13B025]**, to the Indian Institute of Technology Madras, Chennai, is a bona fide record of the research work done by him under the supervision of Dr.Venkatesh Balasubramanian which awards the degree of **Master of Technology and Bachelor of Technology.** The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Dr. Venkatesh Balasubramanian
Research Guide
Dept. of Engineering Design
IIT-Madras, 600 036

Place: Chennai
Date: 5th May

ACKNOWLEDGEMENTS

I would like to thank my research guide Dr. Venkatesh Balasubramanian, for his supervision and feedback to take this work forward. His insight and ideas helped me broaden my views which led to successful results.

I would also like to thank his PhD student, Mr. Sathish Kumar who guided me throughout the entire project and helped me complete it. His guidance helped me in all the time of research.

Karthikeyan R
5th May 2018

ABSTRACT

The number of road accidents in India have been increasing in the past years. Due to the absence of any large scale study in pedestrian accidents, proper policies preventing such crashes have not been implemented. Road accidents are one of main reasons for physical impairment, preventing them should be one of our top priorities. In this study, we identify significant potential risk factors like human factors, environmental factors, road conditions and vehicle condition affecting pedestrian road accidents. The data was acquired from the RADMS database from the Government of Tamil Nadu. We analyze 3147 accidents filed from the year 2014 to 2016 in Tamil Nadu. This study uses Logistic regression to find the relevant risk factors causing injuries. Logistic regression best describes data and explain the relationship between one dependent binary variable and one or more independent variables. Empirical analysis of this study has provided the following major risk factors. Drivers without proper license tend to involve themselves in accidents causing severe injuries. The type of vehicle majorly influences the injury severity. Poor light conditions on the road cause extensive amount of influence in the safety, as it provides difficult situations for both the rider and pedestrian. Bad road conditions too aid in causing grievous injuries. Furthermore, the study describes the explanatory variables along with descriptive statistics and further analyses the results providing suggestions for policy implications.

TABLE OF CONTENTS

LIST OF TABLES

3. LIST OF FIGURES

1. Introduction

Road traffic accidents is a major health issue, due to its alarming increase in number all around the world. World Health Organization says around 50 million people are injured in traffic globally, and warns that road accidents are a leading cause of disability. The economic consequences are also very significant, around 500 billion dollars are being cost due to car crashes alone. Pedestrian road accidents in India have always been at a constant rise every year. The number of pedestrians died in road accidents have increased from 13,894 in 2015 to 15,746 in 2016. Road accidents in India are due to a lot of reasons ranging from poor road designs, lack of proper pedestrian infrastructure, low rates of penalty and even lower rates of conviction. The way of transportation with its rules and regulations have not been fool proof. In a developing country like India, with rapid growth in social and economic development, the density of the road, i.e., the number people using the road both by vehicles and by walk is always at a surge. Due to the increasing population and urbanization, the government has found it hard to bring order in constructing safer roads and walk paths for the pedestrians. Rules are not being adhered by both the drivers and the pedestrians. Figure 1 shows how the accidents along with injuries and fatalities have been constantly increasing every year from 2016 to 2015. Pedestrians are always at risk while on the road, due to absence of protective gears while walking. Since it is a most basic form of physical activity, people pay less attention while walking, and often use electronic devices to text or talk or to listen to music. Riders on the other hand, often tend to drive rashly and faster than the speed limit and their frequent disobedience of the small rules leads up to bitter repercussions. A lot of factors apart from this such as environmental factors, rider's characteristics and pedestrian characteristics play a vital role in these accidents. Some of the factors responsible for road accidents depend on locations. It has been reported by the Transport Research wing of India that, maximum number of accidents (1,88,196) and resultant deaths (62,508) occurred in open area, residential area, market place, pedestrian crossing, etc. When accidents tend to occur in roads where there is no separation or protection for the walk paths for the pedestrians, often it is considered to be the fault of the motor vehicle even though the pedestrian might be liable as they are more vulnerable. Therefore, in this situation of disorder it is quite important to study about the pedestrian related road accidents and the factors causing them. In this paper we study the

road traffic accidents and how the factors are influencing the severity of the injuries using a logistic regression analysis based on the 3417 pedestrian accident reports in the 2014-2016 RADMS database of Tamil Nadu government. The study of risk factors will help develop preventive measures and hence reduce accident rates.

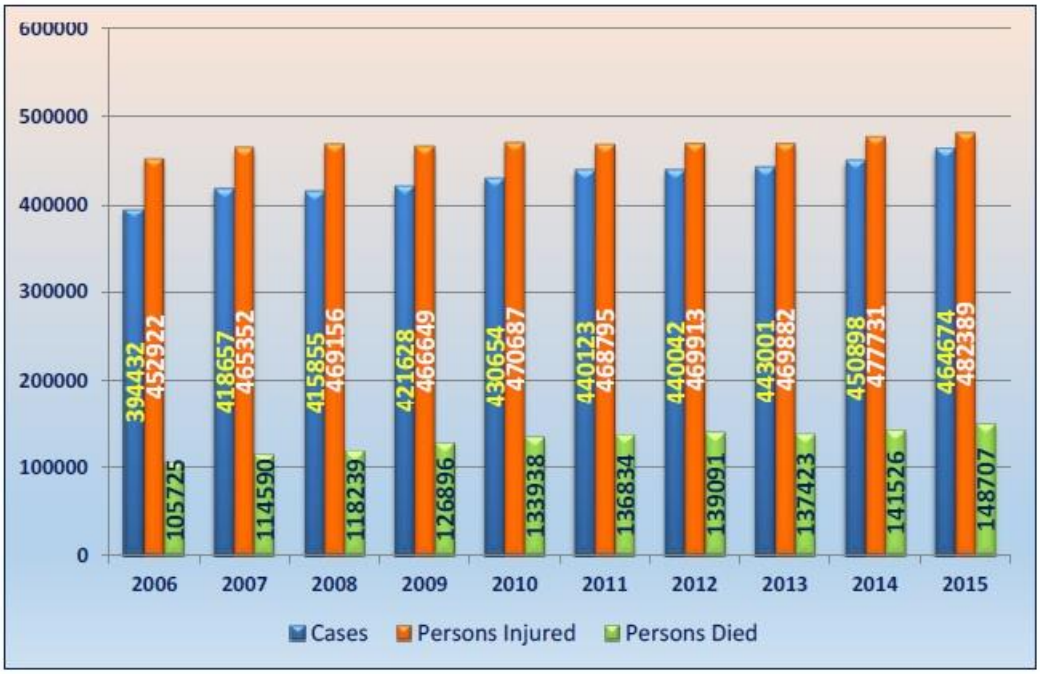

Fig.1 Trend of Road Accidents

2.Study Data and Hypothesis

2.1Data

The Road Traffic Accident data is provided by the Government of Tamil Nadu from the RADMS database. This data is used for further analysis in this paper. The descriptive analysis obtained from this research work provide an example of traffic safety policy making. To assess the pedestrian accident, the dependent variable “severity of injury caused during the accident” was set as Fatal/Grievous Injury--1 and Simple/No Injury--0. Grievous injury is defined as emasculation, or permanent privation of sight of the either eye, or permanent privation of hearing of either ear, privation of any member or joint,

permanent impairing of the powers of any member or joint, permanent disfigurement of head or face, fracture or dislocation of a bone or tooth, or any hurt which endangers life or which causes the sufferer to be during the space of twenty days in severe bodily pain, or unable to follow his ordinary usual pursuits. Fatal injury is an injury resulting in death. Any other injury inflicted which cause bodily pain, disease or infirmity is simple injury. This injury severity along with a lot of other factors were determined by the police officer when the accident occurred, this later was updated in the database. The independent variables which affect the injury severity included driver's and pedestrian's personal information, road condition, light condition, weather condition, type of vehicle, location of the accident and other environmental factors.

Pedestrian's and the driver's personal information such as age, gender, whether the driver has a valid license are considered to be potential risk factors. As young drivers tend to drive rashly and might cause disorder on the road, where as old people might find it hard to take quick turns to steer away from accidents. Hit and run cases might also impact the severity, as riders might want to run away from conviction when involved in accidents with fatal or grievous injuries. Road type is an important factor as the quality of the roads differ from area to area and it definitely varies between busy main roads and the roads at the residential areas. There is a significant difference between highways and urban roads. The quality of the road, the frequency of the roads getting deteriorated, the volume of vehicles it can permit, the reaction time to repair the damaged roads and the quality of the road infrastructure. Many a time urban roads end up getting scraped due to either laying out underwater water pipe lines or electrical lines, even after which roads get laid out only after a considerable amount of time. Availability of proper signal stops or stop signs or police officer at busy junctions to control the traffic enables free flow of vehicles avoiding traffic jams and confusions amongst the riders and the pedestrians. Environmental factors including light condition and weather condition are important factors as well. This is where the infrastructure of the roads play an important role, often during the night or early morning few roads have zero visibility due to no street lamps and the rider solely depends on his vehicle's headlights. This is a probable cause for accidents to occur, especially with pedestrians involved who has no source of light. Harsh weather conditions cause potential risk as both the riders find it hard to detect obstacles and maneuver them during such

situations. The drivers often have less control during such conditions due to low visibility and unfavorable road conditions. Crash factors, i.e., how the collision occurred might also influence the severity. Whether the main collision was head on, hit from the rear, hit from rear, or directly hit the pedestrian would certainly determine the type of injury. Vehicle types would also be a factor as it enables the riders to ride them differently. Motorcycle riders usually travel a bit reckless than car riders. Also they are more prone to accidents as the protective gears that they wear are not safe enough to avoid major bruises.

2.2Hypotheses

After reviewing and consolidating research papers which studied in similar domain (refer section 5.Reference) along with proper studying of the acquired data on the pedestrian road accidents, the following hypothesis are set.

2.2.1 Severity Analysis

i. Lack of driver license increase injury severity (H1)

ii. Vehicle type has an impact on the severity (H2)

iii. Light condition/ time of accident relates to accident severity (H3)

iv. Condition of the road relates to injury severity (H4)

2.3 Descriptive Statistics

The variables along with its definition and its proportion is listed in Table1. From the acquired data we can clearly see that the majority of the drivers were male (64.1%), and in terms of age the people within range of 18-24 followed by 25-34 were the majority with 24.6% and 23.2%. In terms of infrastructure of the road, majority of the accidents happened where it did not contain any junction control (78%). This indicates that more of the traffic control devices need to be installed more efficiently. For environmental factor around 42% accidents happened during daylight. Major reasons of accident are studied to be speeding (77.4%) and not following rules (12.1%). The density of the vehicles is leaded by motorcycles with 46% followed by LMV with 29%.

Table 1: Descriptive statistics of variable

No.	Variables	Description of Variables	Proportion
1	Collision Type		
	Hit Pedestrian	Hit Pedestrian=1; otherwise=0	0.648
	Head on	Head on=1; otherwise=0	0.083
	Hit from side	Hit from side=1; otherwise=0	0.049
	Hit from rear	Hit from rear=1;otherwise=0	0.023
	Hit object	Hit object=1;otherwise=0	0.013
	Skidding	Skidding = 1; otherwise = 0	0.002
	Others	Others = 1; otherwise = 0	0.180
2	Number of vehicles involved		
	1	1= 1 ; otherwise = 0	0.90
	More than 1	More than 1 = 1 ; otherwise = 0	0.10
3	Number of drivers involved		
	1	1 = 1 ; otherwise = 0	0.915
	More than 1	More than 1 = 1 ; otherwise = 0	0.084
4	Number of pedestrians involved		
	1	1=1;otherwise = 0	0.978
	More than 1	More than 1 = 1; otherwise = 0	0.021
5	Central Divider		
	Yes	Yes = 1 ; otherwise = 0	0.355
	No	No = 1 ; otherwise = 0	0.645
6	Road Category		
	Highway	Highway = 1 ; otherwise = 0	0.686
	Not a Highway	Not a Highway = 1 ; otherwise = 0	0.089
7	Road Condition		

	Good	Good = 1 ; otherwise = 0	0.998
	Poor	Poor = 1 ; otherwise = 0	0.001
8	Light Condition		
	Darkness	Darkness = 1 ; otherwise = 0	0.114
	Streetlight	Streetlight = 1 ; otherwise = 0	0.225
	Daylight	Daylight = 1 ; otherwise = 0	0.428
9	Speed Limit		
	30	30 = 1; otherwise = 0	0.0008
	35	35 = 1 ; otherwise = 0	0.004
	40	40 = 1 ; otherwise = 0	0.983
	50	50 = 1; otherwise = 0	0.010
10	Traffic Movement		
	One-Way	One-Way = 1 ; otherwise = 0	0.014
	Two-Way	Two-Way = 1 ; otherwise = 0	0.985
11	No. of lanes		
	1	1 = 1; otherwise = 0	0.837
	2	2 = 1 ; otherwise = 0	0.151
	Greater than 2	Greater than 2 = 1 ; otherwise = 0	0.011
12	Road Works		
	Yes	Yes = 1 ; otherwise = 0	0.020
	No	No = 1 ; otherwise = 0	0.980
13	Accident Cause		
	Alcohol abuse	Alcohol abuse = 1 ; otherwise = 0	0.002
	Animal involved in accident	Animal involved in accident = 1 ; otherwise = 0	0.0008
	Changing lane without due care	Changing lane without due care = 1 ; otherwise=0	0.019
	Dangerous overtaking	Dangerous overtaking = 1; otherwise = 0	0.020

	Driving against flow of traffic	Driving against flow of traffic = 1; otherwise = 0	0.016
	High speed	High speed = 1; otherwise = 0	0.774
	Inattentive turn	Inattentive turn =1; otherwise = 0	0.008
	Injured in accident	Injured in accident =1; otherwise = 0	0.023
	Non-respect of rules	Non-respect of rules = 1; otherwise = 0	0.121
14	Weather Condition		
	Cloudy	Cloudy = 1; otherwise = 0	0.010
	Fine	Fine = 1; otherwise = 0	0.985
	Rainy	Rainy = 1; otherwise = 0	0.003
15	License Type		
	Full	Full = 1; otherwise = 0	0.718
	No License	No License = 1; otherwise = 0	0.159
16	Day		
	Sunday	Sunday = 1; otherwise = 0	0.141
	Monday	Monday =1 ; otherwise = 0	0.149
	Tuesday	Tuesday = 1 ; otherwise = 0	0.137
	Wednesday	Wednesday = 1; otherwise = 0	0.144
	Thursday	Thursday = 1; otherwise = 0	0.139
	Friday	Friday = 1; otherwise = 0	0.146
	Saturday	Saturday =1; otherwise = 0	0.140
17	Driver Gender		
	Male	Male=1; otherwise = 0	0.641
	Female	Female=1; otherwise = 0	0.013
18	Pedestrian Gender		
	Male	Male = 1; otherwise = 0	0.505
	Female	Female=1; otherwise = 0	0.234
19	Season		

	Autumn	Autumn=1; otherwise = 0	0.257
	Spring	Spring=1; otherwise = 0	0.259
	Summer	Summer=1; otherwise = 0	0.259
	Winter	Winter=1; otherwise = 0	0.224
20	Zone		
	East	East=1; otherwise = 0	0.177
	West	West=1; otherwise = 0	0.207
	North	North = 1; otherwise = 0	0.174
	South	South = 1; otherwise = 0	0.439
21	Time		
	Early Morning	Early Morning=1; otherwise = 0	0.098
	Morning	Morning =1; otherwise = 0	0.206
	Noon	Noon=1; otherwise = 0	0.179
	Evening	Evening=1; otherwise = 0	0.211
	Night	Night=1; otherwise = 0	0.243
	Midnight	Midnight=1; otherwise = 0	0.061
22	Hit and Run		
	Yes	Yes=1; otherwise = 0	0.150
	No	No=1; otherwise = 0	0.765
23	Junction Control		
	Control	Control=1; otherwise = 0	0.78
	No Control	No Control =1; otherwise = 0	0.197
24	Driver age		
	<18	<18 = 1; otherwise = 0	0.021
	18-24	18-24 = 1; otherwise = 0	0.246
	25-34	25-34 = 1; otherwise = 0	0.232
	35-44	35-44 =1; otherwise = 0	0.130

	45-54	45-54 = 1; otherwise = 0	0.075
	55-64	55-64 =1; otherwise = 0	0.032
	>65	>65 =1; otherwise = 0	0.007
25	Pedestrian age		
	<18	<18 = 1; otherwise = 0	0.057
	18-24	18-24 = 1; otherwise = 0	0.044
	25-34	25-34 = 1; otherwise = 0	0.214
	35-44	35-44 =1; otherwise = 0	0.126
	45-54	45-54 = 1; otherwise = 0	0.144
	55-64	55-64 =1; otherwise = 0	0.153
	>65	>65 =1; otherwise = 0	0.124
26	Vehicle type		
	Bus	Bus =1 ; otherwise = 0	0.038
	HGV	HGV =1; otherwise = 0	0.033
	Human powered vehicle	Human powered vehicles = 1; otherwise = 0	0.0008
	LMV	LMV =1; otherwise = 0	0.29
	Motorcycle	Motorcycle=1; otherwise = 0	0.46

3.Methodology

3.1. Logistic Regression

In our study the output variable, injury severity, is a binary or dichotomous variable. Logistic regression uses an equation for representation, much like linear regression. Input

variables are combined linearly using weights or coefficient values to predict an output value y. In Logistic regression the output value being modelled is binary values and not a numeric value. Below is an example of logistic regression equation:

$$Y = \frac{e^{(b0+b1*x)}}{(1+e^{(b0+b1*x)})}$$

Where y is the predicted output, b0 is the bias or intercept term and b1 is the coefficient for the single input value (x).

Logistic regression is a predictive analysis just like other regression analysis. The model is used to describe data and explain the relationship between one dependent binary or dichotomous variable and one or more ordinal, interval, nominal or ratio level independent variables. Therefore, logistic regression is the best suitable technique in our case, since it is developed to predict binary dependent variable as a function of predictor variables. We use this model to analyze risk factors in the accident analysis. The probability of an event of interest (i.e., severity of injury) is expressed as:

$$Prob(y = 1|x) = \frac{e^{(x'\beta)}}{(1+e^{(x'\beta)})} = \Delta(x'\beta)$$

Where Δ denotes the cumulative distribution function. β is the vector of parameters to be estimated and x is the corresponding design matrix with vectors of independent variables. When an independent variable x_i increases by one unit when all other variables are kept constant, the odds increase or decrease by a factor of $e^{(\beta i)}$, this factor is called the odds ratio(OR), it ranges from 0 to positive infinity. An odds ratio is a measure of association between an exposure and an outcome. The OR represents the odds that an outcome will occur in the presence of an exposure, compared to the odds of the outcome that occurs in the absence of the outcome. The ratio here indicates that whether the odds of an event of interest increases (when OR >1) or decreases (when OR<1) when the corresponding independent variable increases by a 1 unit.

Table 2: Analyzing severity risk in pedestrian – motor accidents

No.	Variables	Odds Ratio	p-value	Coefficient	5%CI	95% CI
1	Collision Type (Relative to Head on)					
	Hit Pedestrian	1.55	0.008	0.43	0.11	0.76
	Hit from side	1.075	0.767	0.072	-0.41	0.55
	Hit from rear	1.507	0.177	0.41	-0.18	1.007
	Hit object	1.55	0.260	0.43	-0.32	1.20
	Skidding	0.175	0.177	-1.73	-4.2	0.78
	Others	0.75	0.346	-0.28	-0.86	0.30
2	Number of vehicles involved					
	More than 1	2.73	0.007	1.007	0.27	1.74
3	Number of drivers involved					
	More than 1	0.86	0.69	-0.14	-0.84	0.56
4	Number of pedestrians involved					
	More than 1	1.39	0.19	0.33	-0.16	0.83
5	Central Divider (Relative to No)					
	Yes	0.902	0.334	-0.102	-0.31	0.105
6	Road Category (Relative to Highway)					
	Not a Highway	0.883	0.404	-0.123	-0.41	0.166
7	Road Condition (Relative to Good)					
	Poor	4.954	0.181	1.6	-0.74	3.94
8	Light Condition (Relative to Darkness)					
	Streetlight	1.001	0.990	0.001	-0.29	0.294
	Daylight	1.125	0.433	0.11	-0.17	0.41
9	Traffic Movement (Relative to One-Way)					

	Two-Way	0.968	0.920	-0.03	-0.66	0.59
10	No. of lanes (Relative to 1)					
	2	1.697	0.155	0.52	-0.20	1.25
	Greater than 2	0.173	0.004	-1.74	-2.95	-0.5
11	Road Works (Relative to No)					
	Yes	3.306	0.00004	1.19	0.62	1.76
12	Speed Limit (relative to 30)					
	35	0.64	0.75	-0.43	-3.23	2.35
	40	0.34	0.4	-1.07	-3.62	1.46
	50	0.138	0.152	-1.97	-4.68	0.72
13	Accident Cause (Relative to Alcohol Abuse)					
	Animal involved in accident	1.52	0.77	0.41	-2.41	3.25
	Changing lane without due care	0.76	0.74	-0.26	-1.87	1.34
	Dangerous overtaking	1.71	0.49	0.53	-1.00	2.08
	Driving against flow of traffic	2.02	0.37	0.70	-0.85	2.26
	High speed	1.32	0.70	0.27	-1.16	1.72
	Inattentive turn	0.82	0.82	-0.19	-1.92	1.53
	Injured in accident	0.16	0.05	-1.78	-3.57	0.004
	Non-respect of rules	25.94	0.00002	3.25	1.75	4.76
14	Weather Condition (Relative to Cloudy)					
	Fine	1.05	0.89	0.05	-0.74	0.85
	Rainy	0.087	0.021	-2.43	-4.51	-0.35
15	License Type (Relative to Full)					
	No License	1.34	0.01	0.29	0.07	0.52
16	Day (Relative to Friday)					
	Sunday	0.709	0.022	-0.34	-0.63	-0.04

	Monday	0.96	0.78	-0.039	-0.322	0.24
	Tuesday	0.86	0.31	-0.14	-0.43	0.14
	Wednesday	0.82	0.20	-0.18	-0.47	0.10
	Thursday	0.75	0.064	-0.27	-0.56	0.016
	Saturday	0.76	0.074	-0.26	-0.55	-0.02
17	Driver Gender (Relative to Female)					
	Male	0.88	0.71	-0.12	-0.77	0.52
18	Pedestrian Gender (Relative to Female)					
	Male	0.89	0.25	-0.11	-0.31	0.08
19	Season (Relative to Autumn)					
	Spring	1.69	0.000008	0.52	0.29	0.76
	Summer	1.66	0.00001	0.51	0.28	0.74
	Winter	1.74	0.000001	0.55	0.32	0.78
20	Zone (Relative to East)					
	West	0.83	0.22	-0.18	-0.47	0.11
	North	0.77	0.11	-0.24	-0.55	0.059
	South	0.74	0.02	-0.29	-0.55	-0.03
21	Time (Relative to Early Morning)					
	Morning (07:00 – 11:00)	1.18	0.31	0.16	-0.15	0.49
	Noon (11:00 – 15:00)	1.13	0.44	0.13	-0.20	0.46
	Evening (15:00 – 19:00)	0.95	0.78	-0.04	-0.35	0.27
	Night (19:00 – 23:00)	1.06	0.71	0.06	-0.26	0.38
	Midnight (23:00 – 03:00)	1.21	0.36	0.19	-0.22	0.62
22	Hit and Run (Relative to No)					
	Yes	0.44	0.000001	-0.81	-1.15	-0.48

23	Junction Control (Relative to Control)					
	No Control	2.52	0.0001	0.92	0.44	1.40
24	Driver age (Relative to <18)					
	18-24	0.97	0.93	-0.02	-0.55	0.51
	25-34	0.95	0.85	-0.05	-0.59	0.49
	35-44	0.99	0.99	-0.0005	-0.56	0.56
	45-54	1.07	0.80	0.07	-0.52	0.67
	55-64	1.13	0.7	0.12	-0.55	0.81
	>65	1.01	0.97	0.013	-1.02	1.04
25	Pedestrian age (Relative to <18)					
	18-24	0.63	0.06	-0.45	-0.93	0.019
	25-34	0.83	0.36	-0.18	-0.57	0.21
	35-44	0.64	0.02	-0.43	-0.80	-0.06
	45-54	0.58	0.003	-0.53	-0.89	-0.17
	55-64	0.68	0.04	-0.37	-0.73	-0.013
	>65	0.55	0.001	-0.58	-0.96	-0.21
26	Vehicle type (Relative to Bus)					
	HGV	1.05	0.86	0.05	-0.57	0.68
	Human powered vehicle	11.32	0.05	2.42	-0.04	4.90
	LMV	1.79	0.013	0.58	0.11	1.05
	Motorcycle	1.71	0.02	0.54	0.07	1.00

Log likelihood = -1898.811 Confidence interval = 95%

3.2Analyzing injury severity in pedestrian accidents

A detailed analysis of the risk factors affecting the injury severity is shown in Table 2. In terms of light condition, crashes during daylight caused severe injuries(OR=1.125). License defaulters(OR=1.34) increased the probability of sever/grievous injuries. Like previously discussed about the infrastructure of the roads, the injuries were severe due to the presence of road work(OR=3.30) and when the condition of the road was bad(OR=4.95). Also when the riders tend to dangerously overtake(OR=1.71), drive against the flow of traffic(OR=2.02), majorly don't follow the rules of traffic(OR=25.94) and when animals are involved in crashes(OR=1.52) it is more likely to experience grievous injuries. As for the time of day when the accident occurred, during midnight the accidents were severe(OR=1.21) probably due to speeding excessively on empty roads, this was followed by accidents in morning(OR=1.18) and noon(OR=1.13). Although during evening, due to busy traffic the crowded roads could not aid in any severe accidents(OR=0.95). Accidents happened during all the seasons caused severe injuries the greatest being winter(OR=1.74) followed by Spring (OR=1.69) and Summer(OR=1.66) The age of the pedestrian did not yield any specific risk factors, but the age of the drivers did. Drivers with the age of 55-64 (OR=1.13) and 45-54(OR=1.07) tend to be involved in accidents with severe injuries. Furthermore, severe injuries were found more frequently in pedestrian accidents when crashes involved human powered vehicles(OR=11.32), LMV (OR=1.79) and motorcycles (OR=1.71).

3.3ROC Curve

ROC curve abbreviated as Receiver Operation Characteristic curve is a graphical representation of the trade-off between false positives and false negative rates for every possible cut-off. The false positive rate(Specificity) is represented on the X-axis and the true positive rate(Sensitivity) is represented on the Y-axis. The accuracy of the test i.e., the ability of the test to correctly classify cases with a certain condition and cases without the condition, is measured by the area under the ROC curve. Our study has an area under the curve of 0.77.

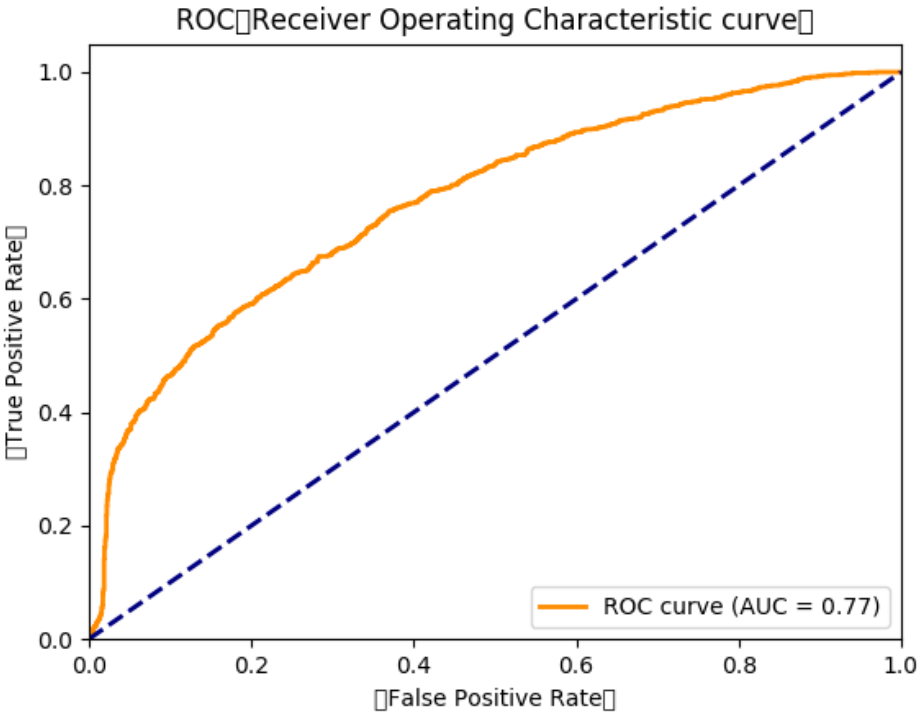

Fig.2 ROC curve

4.Discussion

The results obtained from this study help us asses the set hypothesis. Vehicles hit from rear, hit on objects and hit pedestrians have similar impact on the injury severity, more towards fatal. But skidding has led to simple accidents only. Coming to infrastructure of the road, poor road conditions (H4) is one of the major contributions to grievous injuries. Poor road conditions make it risky for the riders to travel safely, especially the motorcycle riders, they have high chances of losing balance and end up crashing causing grievous injuries either to pedestrians or to themselves. The quality of the roads must be improved in order to travel safely and reduce deterioration. Furthermore, road facilities such as traffic signals and stop signs affect the probability of these accidents.

In terms of human factors, non-respect of rules is one of the main reasons for grievous/fatal injuries in accidents. Riders without any proper license(H1) adds to the risk factor of inflicting grievous injuries. Most of the underage drivers are being easily escaped. Reasons being high vehicle density on the roads, lack of governance – bribery and lack of strict norm and punishments. This again should be properly taken care of with a fool proof system. In addition, no correlation is found between the age of the pedestrian and injury severity.

Environmental factors such as the time of occurrence of the accident is a crucial factor. Most of the accidents occur at midnight (23:00 – 03:00) and morning (07:00 – 11:00)(H3), midnight because the roads are mostly free and people tend to over-speed than the usual speed limit and risk in losing control of the vehicle, cases of drunk & drive are also most likely at this time period. Availability of police officers at night are also less, aiding the defaulters to run away easily. Morning being the start of the for many schools and organization, it is time period where traffic begins to start and continue for a while, people being in hurry to tend to drive rashly in congested and disorganized traffic.

Under vehicle factors, human powered vehicles mainly contribute towards high severity in injury(H2), mostly on the receiving end. Main reasons being no safety features like ABS or airbags or any protective gears like helmet. Therefore, for any accident, the severity of the injury is steep. LMV and motorcycle also have considerable amount of influence in the severity. The Government should take steps in ensuring safety features being built at the stage of design, manufacture, usage, operation and maintenance of both motorized and non-motorized vehicles in par with international standard and practices to avoid any trivial malfunction and to minimize adverse safety and environmental effects of vehicle operation on road users and infrastructure.

Upon analysis, certain policy suggestions have been formulated. Firstly, proper regulation should be followed in making sure the driver has proper essential license to drive the corresponding vehicle. The entire process of earning a driver's license itself is very easily rigged. Not much training is required to get the license, the tests required to pass in order to attain the license should be made tougher. Second, roads must be maintained properly in order to facilitate the riders to drive at any time. Third, proper infrastructure of the road

must be maintained properly. The roads that are being dug out for other purposes should be re-laid as soon as the work is complete, the quality of the roads should be improved, traffic signals and stop signs should be put in necessary places, proper working streetlamps must be installed in places that require them so that people could safely travel in dark, adequate pathways with barriers should be laid in order for the pedestrians to walk. Fourth, proper governance and strict norms and punishments to be established in order to reduce defaulters at road causing accident. The penalty for drunk driving, speeding, driving without proper license and other violation of the rules must be strengthened. Fifth, the government should take steps to ensure proper safety features are built in both motorized and non-motorized vehicles in order to minimize any order of malfunction and causing hindrance to the environment.

India is a country with growing socio economic changes, population and rapid motorization. Traffic analysis on these ever changing roads is a challenging task. Lack of good and sufficient data is common. Methods to store appropriate data is becoming essential in order to conduct proper study and research work to find new solutions. Data provided should also be consistent apart from being accurate, failure of which affect the efficiency of the research work done. Only when good quality of data is collected scaling up the study would be easier. This study provides an overview of risk factors which were significant in pedestrian related crashes influencing the injury severity. Further wide range of data would help scale up and explore different aspects of the accidents in various demographs and prevent them.

5. Reference

1. Analyzing fault and severity in pedestrian–motor vehicle accidents in China. Guangnan Zhang, Kelvin K.W. Yau , Xun Zhang. *Accident Analysis and Prevention*
2. A logistic model of the effects of roadway, environmental, vehicle, crash and driver characteristics on hit-and-run crashes. Richard Tay, Shakil Mohammad Rifaat, Hoong Chor Chin. *Accident Analysis and Prevention*

3. Road Accidents in Inda – 2016. Government of India Ministry of Road Transport & Highways Transport Research Wing.

4. Using data mining techniques to road safety improvement in Spanish roads. Luis Martín, Leticia Baena, Laura Garach, Griselda López and Juan de Oña

5. The role of human factor in incidence and severity of road crashes based on the CART and LR regression: a data mining approach. Alireza Pakgohar a, Reza Sigari Tabrizi , Mohadeseh Khalili , Alireza Esmaeili

6. A data mining approach to investigate the factors influencing the crash severity of motorcycle pillion passengers. Ali Tavakoli Kashani, Rahim Rabieyan, Mohammad Mehdi Besharati