CROSS/ ROADS\
LINKING MOBILITY SOLUTIONS

IRF WORLD ROAD MEETING 2017
/ 14-17 NOVEMBER / DELHI / INDIA /

IDENTIFICATION OF FACTORS INFLUENCING INJURY SEVERITY OF MOTORIZED TWO- WHEELER CRASHES IN PATNA

Speaker Name- Mr. Sachin Kumar Gupta
Designation- Research Scholar (Civil Engineering Department)
Organization- MNNIT, Allahabad
Email Address- sachin0152@gmail.com

PRESENTATION OUTLINE

- Introduction
- Objective of the study
- Data Mining : An Overview
- Methodology
- Results
- Conclusions
- References

INTRODUCTION

Global Road Safety Scenario-

Global Status Report on Road Safety, (WHO, 2015)

•1.25 million people are killed each year on the world's road.

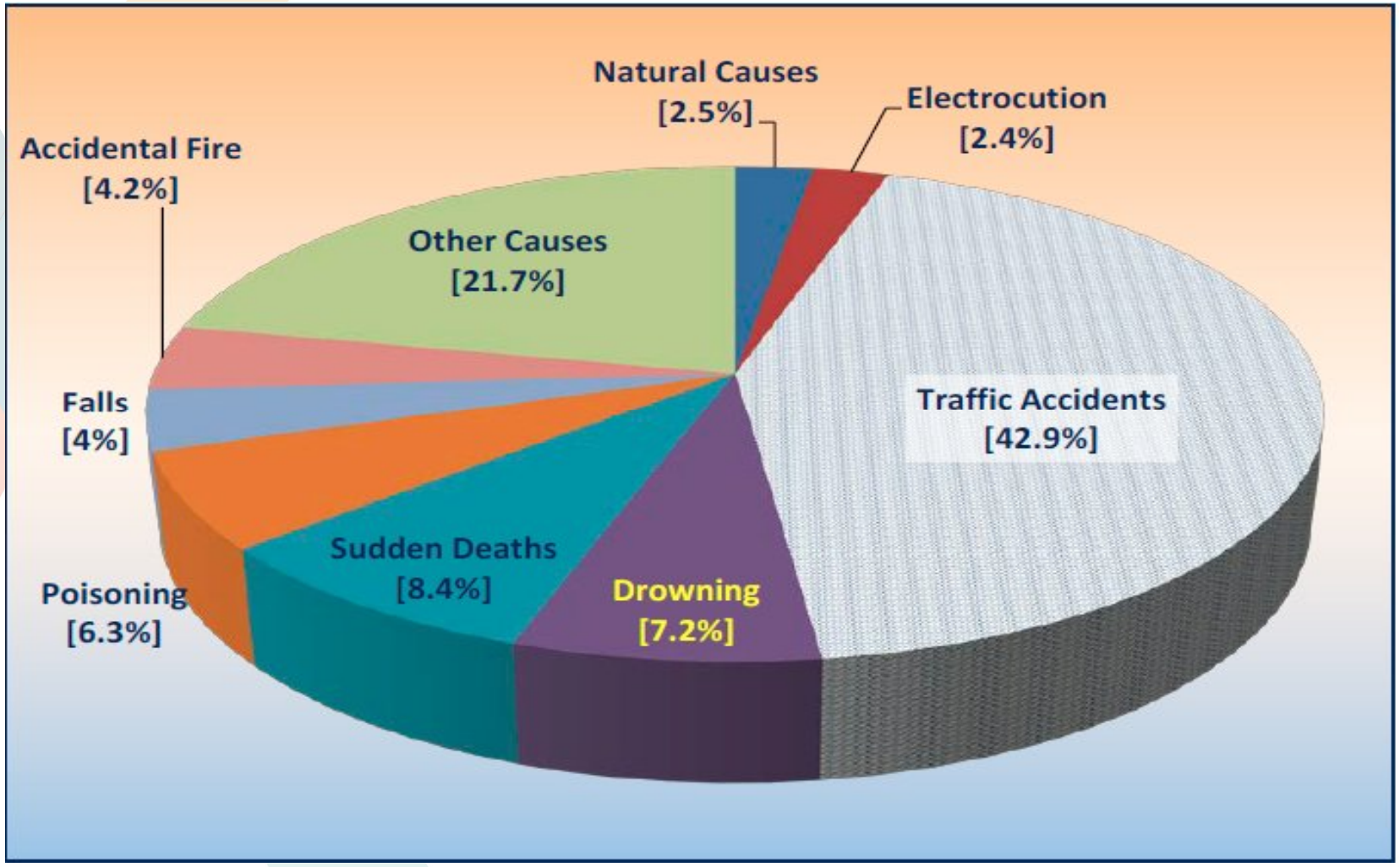
•Leading cause of the death among15-29 yr.

•74% of total deaths occurs in middle income countries yet they have only 53% of total vehicles.

•49% Of total deaths are among motorcyclists (23%), pedestrians (22%) and cyclists (4%) (VRUs).

•In 2010 UN General Assembly adopted the resolution calling for 'Decade of action for road safety (2011 - 2020)'.

Indian Road are safe...?



(NCRB, 2015)

Indian Road Safety Scenario-

Road accidents 2015, MoRTH statistics-

- ≈ 5 lacs injuries (1 accident every minute) &1.46 lacs killed (1 death every 3.7 minutes) in 2015.

- 54.10% road accidents victims are in 15- 34 year age.

- Severity index (deaths per 100 crashes) in 2015 was 29.10.

- It is estimated that 3% GDP (100,000 crore) loss every year.

OBJECTIVE OF THE STUDY

- To identify the roadway and roadside factors influencing the severity of crashes in Indian cities like Patna.

- To explore the use of Data Mining technique in road safety particularly in Indian road conditions.

DATA MINING: AN OVERVIEW

- Extraction of knowledge from raw data is called ‘Data Mining’ (Han et al., 2012).
- Represents data in graphical manner so easy to understand process and results for people.
- Has been extensively used in the business, medical, social, environment and astronomical fields.
- Implementation in the traffic safety field is relatively few.

•Parametric techniques such as logistic regression and ordered probit models (Donnell and Conner, 1996: Kockelman and Kweon, 2002: Abdel Aty, 2003: Wang and Kockelman, 2005: Lemp et al., 2011) have been extensively used in severity classification.

• Researchers have used non parametric procedures such as Classification and Regression Tree (CART) and Multivariate Adaptive Regression Splines (MARS).

•Tree based models have been used in severity analysis (kuhnert et al., 2000: Tesema et al., 2005: Chang and Wang, 2006: Kashani and Mohaymany, 2011: Montella, 2011: Griselda et al., 2012: Montella et al., 2012: Kashani et al., 2014).

METHODOLOGY

Past crash data collection from FIRs

Segmentation of stretches

On site investigation & data collection

J48 & Random forest models

Decision tree generation

Identification of factors influencing severity

- Tree based models use ‘Attribute Selection Measures’ such as information gain concept to select the best attribute to spilt the root node.
- Minimum information (bits) needed to classify the attribute.
- Attribute which provides maximum information gain is considered.
- Tree pruning is done to remove the problem of over fitting.

Software- Weka 3.8.1

•Waikato Institute for Knowledge Analysis (Weka 3.8.1)

•Written in Java, developed at University of Waikato, New Zealand

•Free software.

Models Used-

•J48

•Random Forest

Data-

Secondary Data

- Two year (2014-15) was collected for police FIRs available at various police stations.
- Time, victims, vehicle involved, location etc. were noted down.

Primary Data

- Attributes like gap in median, access per km. , presence of school, petrol pump etc. were collected on site.

- Total of 16 attributes were used.

•Data was arranged in different time slots keeping in mind the traffic temporal variations in the city

Time interval	Notation
23:00 - 04:00	T1
04:00 – 06:00	T2
06:00 – 08:00	T3
08:00 – 12:00	T4
12:00 – 16:00	T5
16:00 – 20:00	T6
20:00 – 23:00	T7

Table 1 Representing time slots

Severity variations with respect to time..

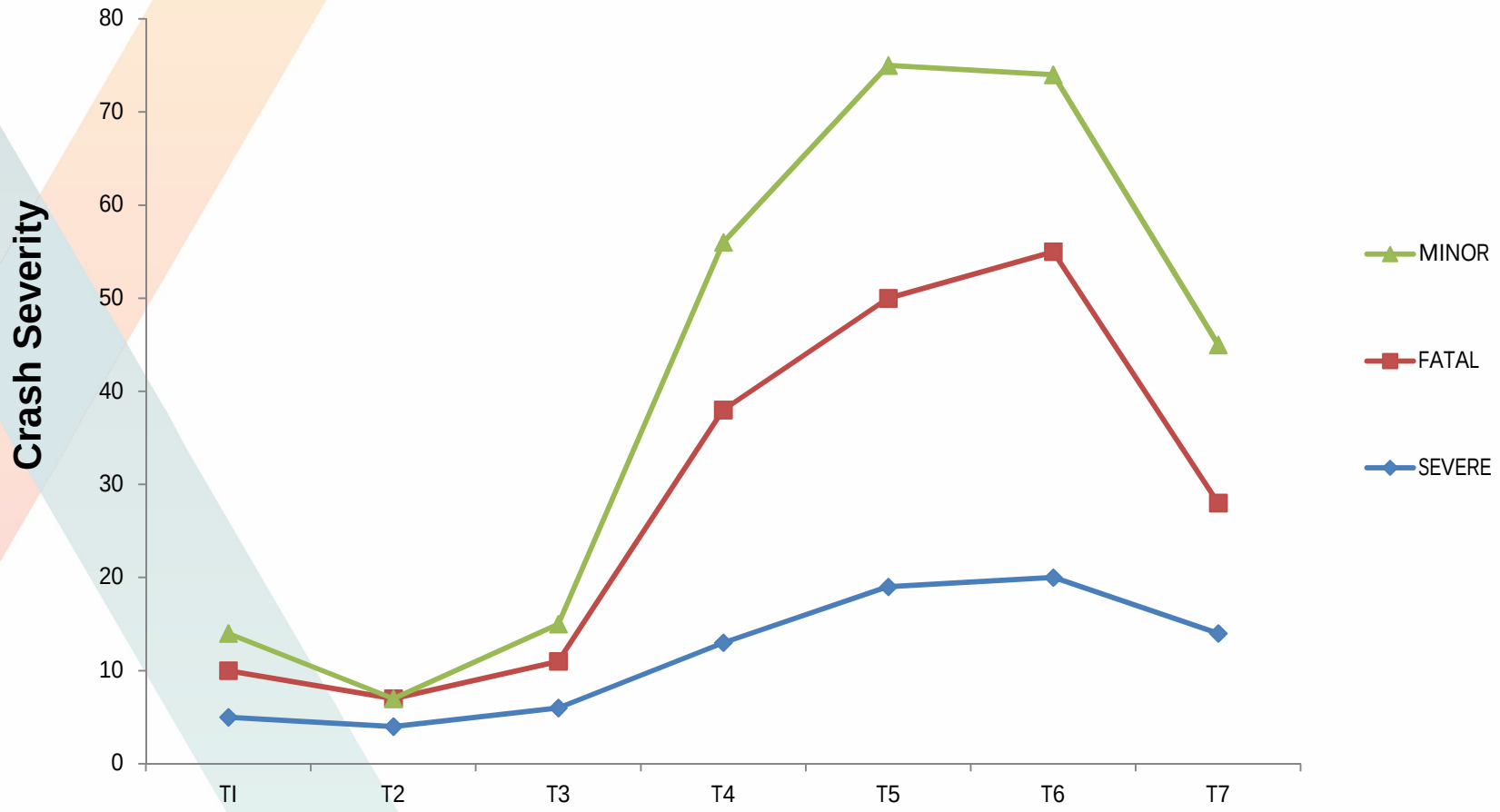


Fig. 1 Representing Severity variations v/s time slots

RESULTS

- Total 219 data point was used for the analysis.
- 80% split was used i.e. 80% for training and 20% data for testing purpose.
- Weka's default parameter's value were used.

Overall Accuracy of Models

S. No.	Model Used	Kappa Statistics	Accuracy in Percentage
1	J48	0.3189	54.545
2	Random Forest	0.3704	59.09

Confusion Matrix Representation-

Predicted class

Actual Class

	Yes	No	Total
Yes	TP	FN	P
No	FP	TN	N
Total	P’	N’	

- J48 Confusion Matrix

a	b	c
6	3	5
5	10	3
2	2	8

a = severe
b = Minor
c = fatal

J48: Detailed accuracy by class

TP Rate	Precision	ROC Area	Class
0.429	0.462	0.692	Severe
0.556	0.667	0.703	Minor
0.667	0.500	0.725	Fatal

Table 2 Representing values of different accuracy measures by class

Random Forest

- Confusion Matrix

a	b	c
9	13	2
2	13	3
4	4	4

- Detailed Accuracy by Class

TP Rate	Precision	ROC area	Class
0.643	0.600	0.649	Severe
0.722	0.650	0.792	Minor
0.333	0.444	0.628	Fatal

CONCLUSIONS

- J48 model was identified as good in predicting fatal crashes having high TP rate as 0.667.

- Random forest model was identified as good in predicting the minor crashes and sever crashes in the city having TP rate 0.722 and 0.643 respectively.

- Kappa values of J48 and Random forest model were observed as 0.32 and 0.37 which falls in fair agreement range (Landis and Koch, 1977).

- No. of access/km was found to be single most important attribute in predicting crash severity.

- Time of the day, land use, median openings, shoulder conditions, parked vehicles, warning signs, length of the stretch, street lights and turning traffic were found to be significant.

REFERENCES

•Aty, A. M. (2003). *Analysis of driver injury severity levels at multiple locations using ordered probit models*. Journal of Safety Research 34 (2003) 597– 603.

•Beshah T., Hill S., (2010). *Mining road traffic accident data to improve road safety: Role of road related crash factors on accidents severity in Ethiopia*. AAAI Spring Symposium .

•Chang, L.Y. & Wang, W.H. (2006). *Analysis of traffic injury severity: An application of non-parametric classification tree techniques.* Accident Analysis and Prevention 38, 1019–1027 .

•Donnell' O, C. J., Conner, H. D. (1996). *Predicting the Severity of Motor Vehicle Accident Injuries using models of Ordered Multiple Choice*. Accident Analysis and Prevention Vol. 28, No. 6, pp. 739-753, 1996

•Global Status Report on Road Safety (2015). WHO

•Griselda, L., Juan, O. D. & Joaquin, A. (2012). *Using Decision Trees to extract Decision Rules from Police Reports on Road Accidents*. SIIV - 5th International Congress - Sustainability of Road Infrastructures. Procedia - Social and Behavioral Sciences 53 (2012) 106 – 114

•Han, J., Kamber, M. and Pei J. (2012). *Data mining: concepts and techniques, 3rd edition*. Morgan Kaufmann, USA.

•Kashani, A. T. & Mohaymany A.S. (2011). *Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models.* Safety Science 49, 1314–1320

•Kashani, A. T., Rabieyen, R. & Besharati, M. M. (2014). *A data mining approach to investigate the factors influencing the crash severity of motorcycle pillion passengers.* Journal of Safety Research 51 (2014) 93–98

•Kuhnert, M. P., Do, A. k., & McClure, R. (2000). *Combining non-parametric models with logistic regression: an application to motor vehicle injury data*. Computational Statistics & Data Analysis 34 (2000) 371{386 }.

•Kockelman, M. K. & Kweon, Y. J. (2002). *Driver injury severity: an application of ordered probit models*. Accident Analysis and Prevention 34 (2002) 313–321.

•Landis, R. J. & Koch G. G. (1977). *The Measurement of Observer Agreement for Categorical Data*. International Biometric Society. Biometrics, Vol. 33, No. 1 (Mar., 1977), pp. 159-174

•Lemp, D. J., Kockelman, M. K. & Unnikrishnan, A. (2011). *Analysis of large truck crash severity using heteroskedastic ordered probit model*. Accident Analysis and Prevention 43 (2011) 370–380

•Ministry of Home Affairs., Accidental death and suicides in India- (2014). National Crime Records bureau

•MoRT&H, *Road Accidents in India* (2015).

•Montella, A. (2011). *Identifying crash contributory factors at urban roundabouts and using association rules to explore their relationships to different crash types*. Accident Analysis and Prevention 43, 1451–1463

Thank You.!!