UGC Approved Journal
© UNIVERSAL RESEARCH REPORTS | REFEREED | PEER REVIEWED
ISSN : 2348 - 5612 | Volume : 05 , Issue : 05 | May 2018

Analysis of Crash data using Machine Learning and Improve design of Road Intersection

[1]Angel Rachana Dasari
Assistant Professor, CVR College of Engineering, Hyderabad

[2]Vamshi Chaitanya Annimalla
Associate Professor, CVR College of Engineering

Abstract:
Road crashes are an issue that affects the whole globe and is now ranked as the ninth leading cause of mortality worldwide. It is also a significant issue in our nation because of the very high number of traffic collisions that occur each year. For this reason, it is very important to analyze crash data and factors influencing it. Telangana state police department records all the crashes occurred in the state. In this project the accident data of year 2019 is collected from Hyderabad traffic police and data is studied for identifying accident patterns using machine learning algorithms and develop improved corridor design addressing the proposed counter measures.

Keywords: Accident Classification, Collision Diagram Development, Machine learning algorithms

Introduction

The primary purpose of this research is to examine the data on road traffic accidents at the level of metropolitan cities using machine learning algorithms and to categorize the accidents according to the kind of collision that occurred. Determine the most pressing concerns about road safety then hold a conversation about possible solutions to those concerns that may be found in the road safety community.

A. Types of Collisions
   1. Rear End collision
   2. Head On collision
   3. Right Angle collision
   4. Side Swipe collision
   5. Pedestrian collision
   6. Parked vehicle collision
   7. Fixed Object collision
   8. Left Turn Collision

B. Machine Learning Algorithms.

Machine learning algorithms used in this project are:
   1. Multinomial Naïve Bayes
   2. Logistic Regression
   3. Random Forest
   4. Perceptron

C. Objectives
   - Identify different types of collision in past 1 year in Hyderabad
   - Use Machine learning algorithm to classify historical crash data of Hyderabad city for year 2019.

UGC Approved Journal
© UNIVERSAL RESEARCH REPORTS | REFEREED | PEER REVIEWED
ISSN : 2348 - 5612 | Volume : 05 , Issue : 05 | May 2018

- Develop collision diagrams and identify crash patterns.
- Study crash patterns and select suitable counter measures.
- Develop improved intersection design addressing the proposed counter measures.

Literature Review

Nidhi (2018) et.al

Accidents involving motor vehicles may be classified into a variety of subtypes, including “rear-end, head-on, and rollover collisions. The state recorded police reports”, also known as FIRs, are the records that include the information that is related to the automobile collisions. In this study, both a priori and naive Bayesian approaches were used to make predictions on the prevalent patterns of vehicular mishaps. This pattern will be helpful to the government or non-governmental organizations (NGOs) in improving safety and taking preventative steps on roads that are high-risk areas for accidents.

Aparna Verma Ashutosh Gupta and Baikonth Nath (2017)

In this research, it was found that the number of accidents that occur on India's roads is a key problem for road safety management. Road Safety Management also identified a few methods that might be taken in “order to reduce the number of accidents that occur on India's roads. The general population received training and participated in an awareness campaign that was planned for them. The number of people who are injured as a result of small” and large accidents is regularly reported, and the number of people who are injured as a result of minor and big accidents is frequently reported and documented for the purpose of furthering India's remarkable progress in road safety.

Jiangfeng Wang, Qian Zhang, Yinhai Wang, Jinxian Weng and Xuedong Yan (2016)

In the investigation of the matched case-control logistic regression model, the only examples of sideswipe accidents that were actually seen were those that occurred on Interstate 5 in the central Puget Sound region of the state of Washington. This is done in order to eliminate the potential for bias that may arise as a result of the large proportion of head-on accidents that are accounted for in the generic models. When the spatial-temporal aspects of the flow of traffic are considered, as well as the short-term variations in the frequency of sideswipe collisions, an exhaustive investigation on the incidence of sideswipe collisions as well as their connection to the flow of traffic on the motorway across lanes and detector placements is carried out. According to the findings, sideswipe incidents are more likely to take place during off-peak hours on stretches of multilane motorways that are both level and straight. In crowded situations, the likelihood of a sideswipe accident is often increased when there is upstream of the event area, there was a high average occupancy, a low average flow, and a low speed variation. In contrast, in uncongested conditions, factors that may reduce the likelihood of sideswipe collisions include a downstream of the crash area there was a high average speed and coefficient of variation in speed, low speed variance and coefficient of variation in occupancy, and a large average absolute difference in speed between neighboring lanes. All of these factors contribute to a lower both the variance in speed and the coefficient of variation in occupancy were measured. In addition, an increase in the average absolute difference in occupancy and speed between upstream and downstream of the site is another factor that contributes to an increase in the number of sideswipe crashes that occur at that location. of the collisions, while an increase in the Contrary to what one would expect, the standard deviation of absolute difference in occupancy and speed has the opposite impact. The model of a collision with a sideswipe that was presented was verified by utilizing the sideswipe

UGC Approved Journal
© UNIVERSAL RESEARCH REPORTS | REFEREED | PEER REVIEWED
ISSN : 2348 - 5612 | Volume : 05 , Issue : 05 | May 2018

collision data from the SR 520 highway. Additionally, A statistical analysis was performed to investigate both the transferability and the stability of the sideswipe collision model that was proposed.

Methodology

A. Data Collection

The gathering of information on the accidents is the initial stage of the accident investigation. The primary responsibility for compiling information on the accidents is with the police. Accident reports reported by motorists themselves are considered secondary data since they are provided by motorists. The following characteristics have to be included in the data that are to be gathered:

1. General: The date, the time, and the people who were involved in the accident, as well as the severity of the event, such as fatal, severe, or minor.
2. Place is the description and specifics of the location where the accident occurred.
3. Information on the car that was involved, including the registration number, a description of the vehicle, details about the loading, and information about any vehicular faults.
4. The nature of the accident, including the specifics of the collision, the damages, the injuries, and the fatalities.
5. The state of the road and the traffic on it, including specifics such as the road's shape, surface qualities, types of traffic, and traffic density, among other things.
6. Details of many probable scenarios that are the primary causes of accidents are provided below as primary causes of accidents.

The following objective calls for the appropriate filing and retrieval of these data that were gathered. The following will explain the goals:

1. The localization of trouble spots with an abnormally high accident rate is something that has to be identified.
2. study of the key accident location's functionality in great detail in order to determine the reasons for accidents.
3. The development of a method that will enable the detection of potential dangers prior to the occurrence of a significant number of incidents.
4. Work is now being done to generate multiple statistical measures of different accident-related features in order to give insight into general patterns, common cause components, driver profiles, and so on. The goal of this work is to provide information that may help prevent future accidents.

The Rachakonda Police Department was the source of the data collection for the Hyderabad (study region). We have compiled the First Information Report (FIR) summaries for each individual case for the year 2019. FIRs are the initial sets of information that are recorded by the police as reported by individuals who were involved in the incident, by eyewitnesses, or by the police official in the absence of any eyewitnesses. Eyewitnesses include anyone who saw the crash or who were engaged in it. A condensed version of the full FIR is referred to as the FIR summary. The FIR summary provides a condensed explanation of the crash facts. It includes information on the time and day that the accident happened, a summary of the event, the cars that were involved, and a general area for where the accident took place. Cases that had been registered but had insufficient information were not included in the study.

UGC Approved Journal
© UNIVERSAL RESEARCH REPORTS | REFEREED | PEER REVIEWED
ISSN : 2348 - 5612 | Volume : 05 , Issue : 05 | May 2018

B. Classification of accidents.

The most important purpose of this undertaking is to classify the crash data into various categories using Machine Learning algorithms. Manually we have classified 400 data into types of collision which can be used for testing and training for machine learning algorithm.

The crash data is classified into 9 categories based on the type of collision.

- Rear End Collision
- Head on Collision
- Side swipe Collision
- Parked Vehicle Collision
- Right Angle Collision
- Left Turn Collision
- Fixed Object Collision
- Pedestrian Collision
- Others

C. Machine Learning Methodology

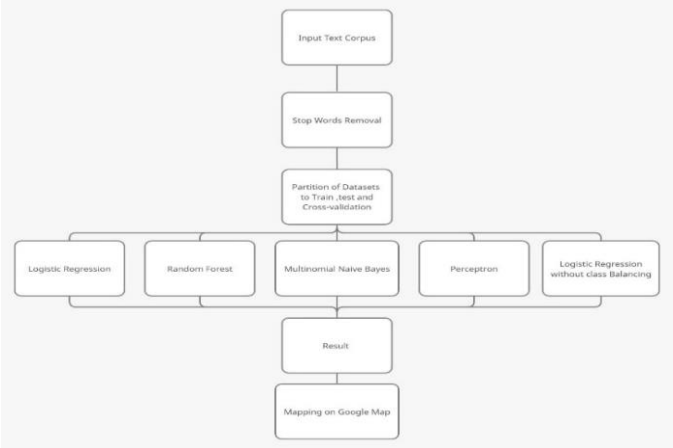


Figure 3.1 Machine Learning Methodology.

Using python language entire code is written and it is executed in Jupyter Notebook. The workflow of the machine learning part is shown in the above Diagram.

D. Cluster Identification using Network screening methods

Crash Frequency Method:

It Ranks locations on basis of average annual number of crashes

Fi = Ci / Ti

Fi = Average crash frequency at location

Ci = Crashes during analysis period,

Ti = Study period (yrs.)

E. Locating crashes in google earthpro

UGC Approved Journal
© UNIVERSAL RESEARCH REPORTS | REFEREED | PEER REVIEWED
ISSN : 2348 - 5612 | Volume : 05 , Issue : 05 | May 2018

- The reports of the collisions are investigated with a focus on the precise area where the collisions took place. The analyst compiles a summary of the collisions according to locations along the roadway.
- Spots are brief stretches of highway that assist in locating "point" places that provide a potential hazard, such as crossroads, bends, or short bridges.
- The data that were gathered are shown on a map of the area, which provides a schematic representation of the crash patterns.

F. Crash pattern identification and Counter measures

- Two criteria are used to categories the reported collisions: the location of the highway and the kind of accident that took place.
- The geometry of the area where the reported collisions took place is taken into consideration for the first step in the categorization process. As a result, accidents are classified according to whether they took place at an intersection, on a portion of highway, or outside of a public roadway. If the accident took place at an intersection, it is recorded within the scope of the category (this case covers driveways, front yards and others).
- To investigate the factors that lead to accidents and to propose preventative actions in vulnerable areas. To assess the present design, determine if it is inadequate, and provide a recommendation for an alternate design for the chosen area.
- Create the updated design in AutoCAD in accordance with the instructions included in IRC: 67-2012 and IRC: 35-2015.

ANALYSIS AND DISCUSSIONS

A. Accuracy of Data Classification

The data classification accuracy greater than 85% is acceptable. The accuracy obtained by the various algorithms for the crash data are given below.

B. Output of Machine Learning Algorithm

Classification of data using machine learning algorithm is done and logistic regression without class balancing algorithm is chosen as it produces 88% accuracy.

Log loss: 0.6072176734515005

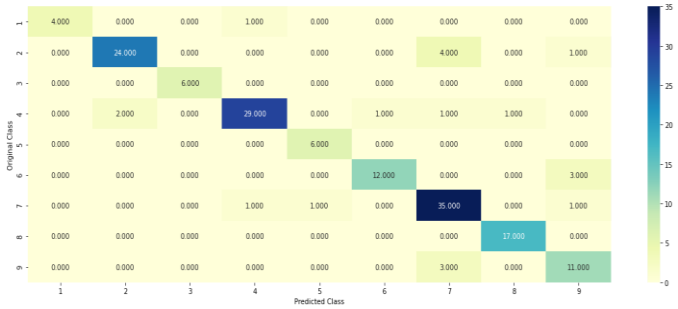
TABLE I. ACCURACY OF ALGORITHMS

S.NO	Algorithm	Accuracy
1	Multinomial Naïve Bayes	85%
2	Logistic Regression	86%
3	Logistic Regression without Class Balancing	88%
4	Random Forest	79%
5	Perceptron	86%

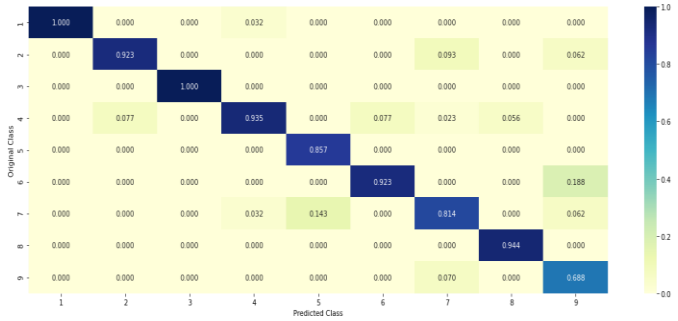
Number of mis-classified points: 0.12195121951219512

UGC Approved Journal
© UNIVERSAL RESEARCH REPORTS | REFEREED | PEER REVIEWED
ISSN : 2348 - 5612 | Volume : 05 , Issue : 05 | May 2018

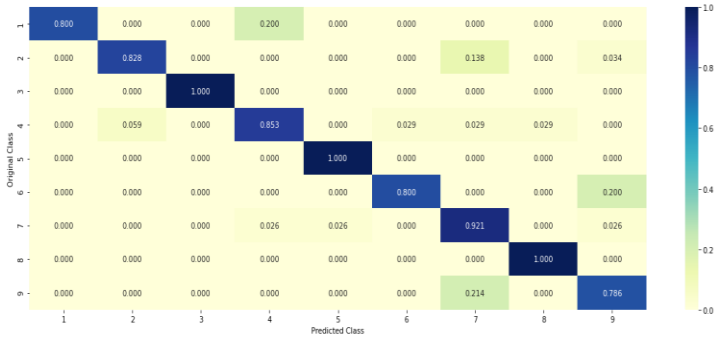
--------- Confusion matrix ----------



--- Precision matrix (Column Sum=1) ----



------ Recall matrix (Row sum=1) -------



Output:

UGC Approved Journal
© UNIVERSAL RESEARCH REPORTS | REFEREED | PEER REVIEWED
ISSN : 2348 - 5612 | Volume : 05 , Issue : 05 | May 2018

Figure 4.1 Output of Logistic Regression without class balancing

The output that is generated by machine learning algorithm of all the 2919 crashes that took place 2019 are comma separated values and later all the values are placed into their respective rows in FIR data.

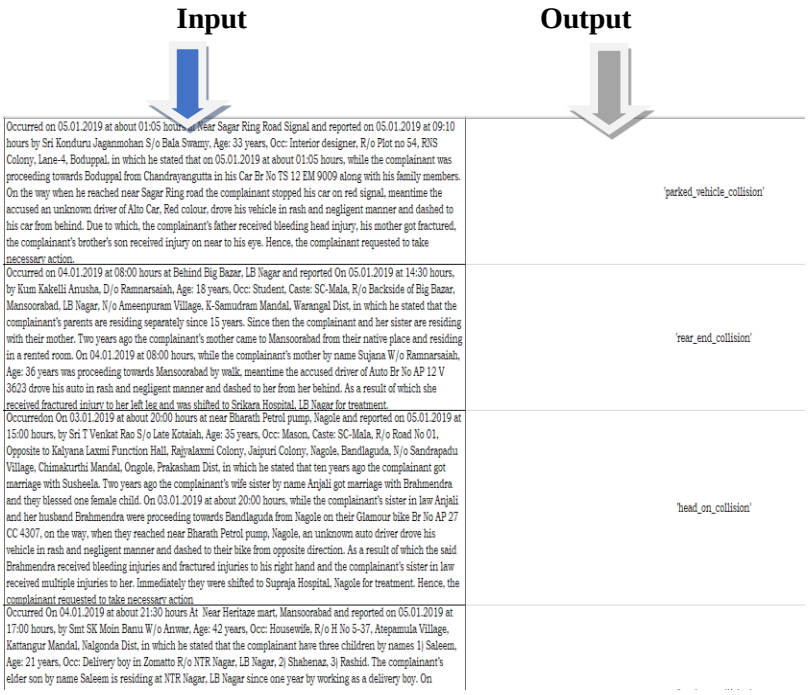


Figure 4.2 Figure Showing input and Output

C. Types of Collision in Hyderabad (2019)

UGC Approved Journal
© UNIVERSAL RESEARCH REPORTS | REFEREED | PEER REVIEWED
ISSN : 2348 - 5612 | Volume : 05 , Issue : 05 | May 2018

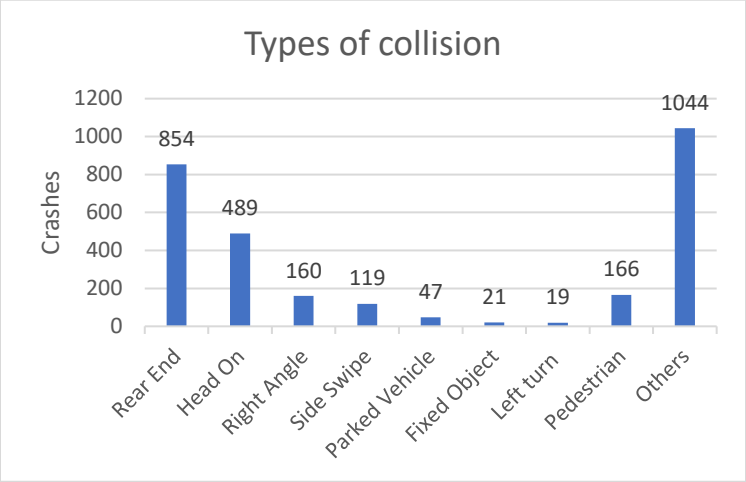


Figure 4.3 Bar graph Showing Types of Collision in Hyderabad (2019)

From the graph we can observe that more Rear End collisions are taking placing in Hyderabad for the year 2019.

D. Identifying Critical locations

Critical locations are identified by Crash frequency Method.

It Ranks locations on basis of average annual number of crashes

Fi = Average crash frequency at location

Frequency= total crashes/study period(yrs.)

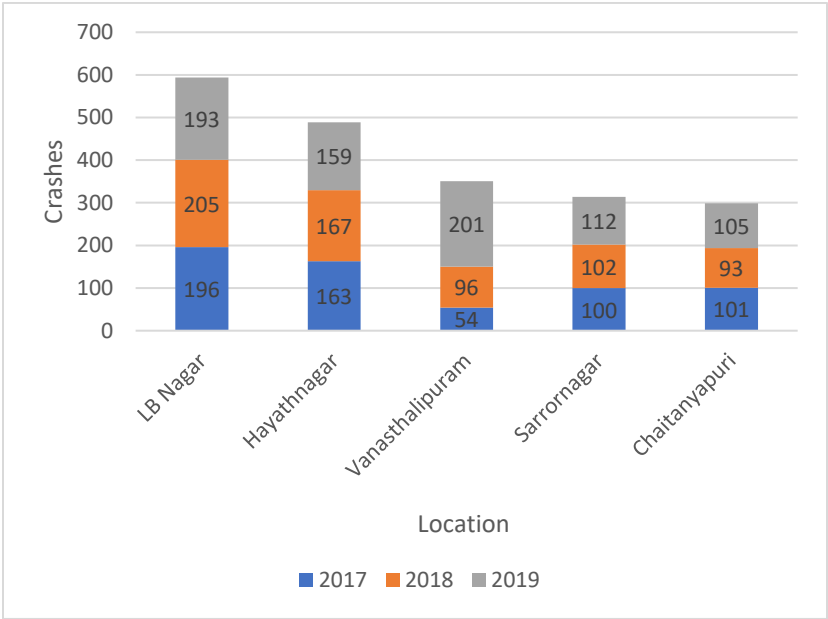


Figure 4.4 Crashes at Critical Locations for year 2017,2018 & 2019.

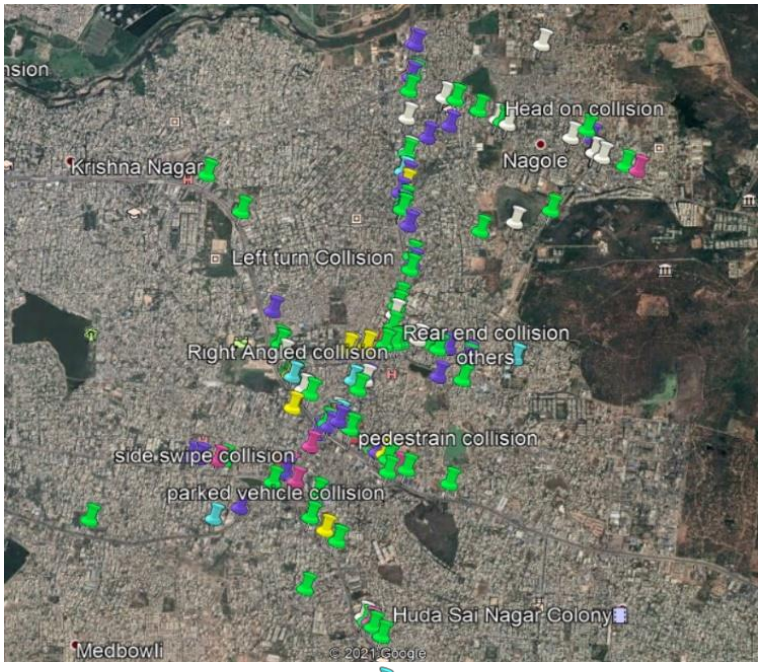
E. Crash pattern identification at LB Nagar

UGC Approved Journal
© UNIVERSAL RESEARCH REPORTS | REFEREED | PEER REVIEWED
ISSN : 2348 - 5612 | Volume : 05 , Issue : 05 | May 2018

Based on the crash data of the past 3 years, LB Nagar region has the greatest number of crashes. So, we have chosen LB Nagar as our Study Area and Located all 193 crashes that took place in the year 2019 from the classified data to find the crash patterns.

TABLE II. COLLISION TYPES AT LB NAGAR

S.NO	Types of Collision	No. of Crashes
1	Rear End Collision	**76**
2	Head on Collision	**27**
3	Pedestrian Collision	**19**
4	Right Angle Collision	**13**
5	Side Swipe Collision	**9**
6	Parked Vehicle Collision	**5**
7	Left Turn Collision	**2**
8	Others	**42**



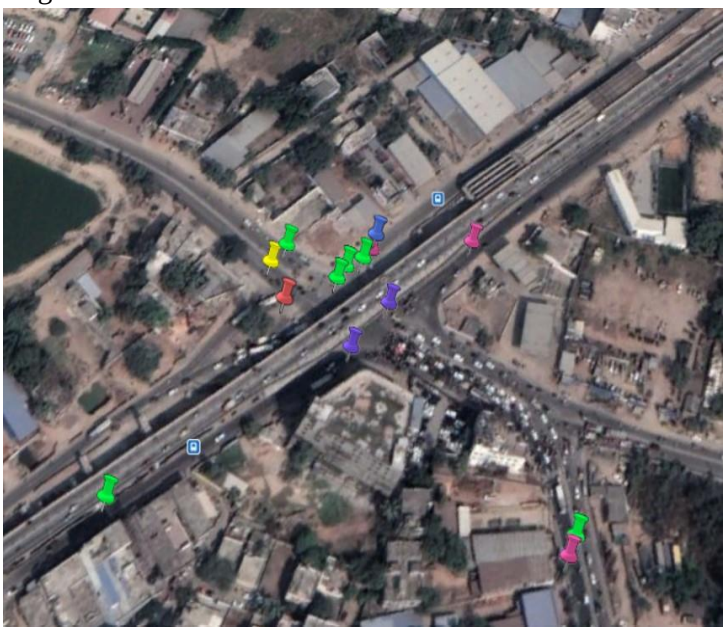
Right angle Collision
Pedestrian Collision
Rear End Collision
Side Swipe Collision
Left turn Collision
Head on Collision
Parked Vehicle Collision
Others

Figure 4.5 Crash patterns at LB Nagar Region

UGC Approved Journal
© UNIVERSAL RESEARCH REPORTS | REFEREED | PEER REVIEWED
ISSN : 2348 - 5612 | Volume : 05 , Issue : 05 | May 2018

F. Crash Pattern at Sagar X Road intersection.

As a greater number of crashes are taking place at Sagar X road intersection we have chosen it as our study area. Crash Patterns at Sagar X road intersection are studied and suitable counter measures are identified.



Right angle Collision
Pedestrian Collision
Rear End Collision
Side Swipe Collision
Left turn Collision
Head on Collision
Parked Vehicle Collision
Others

Conclusion

To reduce the crashes at Sagar X road Intersection the following Countermeasures are proposed

- Channelization should be done in order to separate turning movements.
- Four Channelizing Islands needed to constructed at the turnings in intersection to make a free left turn.
- Lane markings and Arrow markings must be provided as per IRC:35-2015 to provide guidance and information to the drivers and to separate the traffic moving in same direction.
- Right turn lane markings must be provided for the vehicles moving from chintalkunta to L B Nagar and TKR college Road to chintalkunta as huge volume of traffic is moving in this direction.
- Chevron markings (Converging and Diverging) must be provided as per IRC:35-2015 at the channelizing island which separates the traffic moving in their direction.
- Two-way Hazard marker boards must be provided at the constructed channels as per IRC:67-2012.
- Traffic Signal boards must be provided at the signal areas as per IRC:67-2012.

UGC Approved Journal
© UNIVERSAL RESEARCH REPORTS | REFEREED | PEER REVIEWED
ISSN : 2348 - 5612 | Volume : 05 , Issue : 05 | May 2018

- The placement of these sign boards should be done at a suitable distance so that the motorist has a chance to view them and has enough time to react to the situation.

By considering all this countermeasures and conclusions an alternative intersection design for Sagar X Road is proposed using AutoCAD as per IRC:35-2015 and IRC:67-2012.

Figure 5.1 Proposed Intersection Design for Sagar X Road.

References

1. Transferability of Community based collision prediction, American Society of Civil Engineers, 2014.
2. Ms. Nidhi, Analysis of Road Accidents Using Data Mining Techniques.
3. J Stannard Baker, Traffic Accident Investigation Manual, The traffic Institute Northwestern University, 2019.
4. Gopalakrishnan S, A Public Health Perspective of Road Traffic Accidents. J Family Med Prim Care, 2012.
5. Prediction of Crash Severity on Two-Lane based on Fuzzy Classification, American Society of Civil Engineers, 2014.
6. Major Accident factors for Effective Safety Management of Highway Construction Projects, American Society of Civil Engineers, 2013.
7. Road accidents in India, 2009.
8. V F Babkov, Road Condition and traffic safety, MIR Publishers, Moscow, 2019.
9. Analysis of sideswipe collision, American Society of Civil Engineers, 2016.
10. Milan Batista, on the mutual coefficient of restitution in two car collinear collisions, 2006.
11. IRC:35-2015 and IRC:67-2012.