Analysis of Road Traffic Accidents to Improve Safety and Protection

H. Tariq[1], M. M. Iqbal[2], U. Khadam[3], M. A. Al Ghamdi[4]

[1] *Department of Computer Science, The Institute of Management Sciences, Lahore, Pakistan*
[2] *Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan*
[3] *Department of Software Engineering, University of Kotli, Azad Jammu and Kashmir, Pakistan*
[4] *Computer Science Department, Umm Al-Qura University, Makkah City, Saudi Arabia*

[2] munwariq@gmail.com

Abstract- Road safety is one of the global issues in the world during travelling. The road accident causes critical effects in the lives of travellers. Thousands of people lose their lives every day, and hundreds of people are injured in a road accidents. Given the importance of the problem, identifying the causes of traffic accidents is the main objective to reduce traffic accidents. In this paper, we apply data mining algorithms and some statistical analysis of road traffic accidents are performed to find patterns and useful information. Different attributes like weather conditions, road surface, light condition, and severity of casualty are investigated. Apriori algorithm is used to discover association rules, and the ZeroR classifier is used to build the classification model. Our empirical results show that the proposed model could classify road traffic accidents with reasonable accuracy and find out association rules. Which help us to improve safety and prevention from accidents. As seen in the experiential setup, we find casualty severity on the basis of different attributes like road surface, weather condition, light condition, and type of vehicles. Association rules are applied to find out the best rules which prevent accidents and improve safety.

Keywords- Roadway Traffic Accidents, Safety, Association Rules, Classification, Data Mining

I. INTRODUCTION

It is difficult to interpret the maximum use of data, resources are also required because of the growth of many governmental, business, and scientific databases. Nowadays, data is generated in Tera Bytes per second, and it is estimated that the volume of data that is stored in the world's database rapidly grow at the rate of 100% in every twenty months, many organizations generate large amounts of data faster than ever before [1]. These facts show that data is explored more and more. According to the report of (WHO), 1.24 million people lose their lives each year, and 20 to 40 million are seriously injured in road accidents throughout the World [2]. Road accidents are significant reasons for property damage, life loss, health issues, and even permanent disabilities. Many people die due to car accidents around the world. Managing this situation is one of the significant challenges that traffic management authorities face. According to the Association for Safe International Road Travel (ASIRT), nearly 1.25 billion people die in road accidents every single year. Moreover, 20-50 million are injured or disabled.it is the 9th major cause of death around the world and the leading cause of death among young people [3-4]. It costs the US $518 billion globally and costs low and middle-income countries about $65 billion every year. By the year 2030, it is predicted to become the 5th leading cause of death. Effective use of modern techniques can help us manage these situations and reduce the loss that occurs due to accidents; however, accidents can be is predictable, and there is no guarantee that the situation remains the same in the future [5].

Another announced from the Centres for Disease Control and Prevention, that is 100$ billion cost charge by medical care every year. Emergency departments in hospitals suffered from accidents every 10 seconds [6]. Severe types of crashes occur with the rapid growth of volume and traffic speed of road traffic. The traffic safety issue rose worldwide with significant concerns and become one of the key challenging issues. As human beings, all of us want to stay safe and avoid accidents, but there are several reasons for accidents happening like the driver, road and vehicle, etc [7-8]. The increase in traffic on the roadway also grows the number of accidents, which is a problem that should be solved, and it is not suited for smart cities. This fundamental problem does not only occur in many applications, such as marketing, technology, finance, military, linguistics, and biology but also in the pre-processing of basic data for each application.

Traffic accidents have a massive impact on the economy due to the cause of deaths and injuries.

Currently, researchers pay more attention to decisive factors that affect significantly traffic accidents. Therefore it is crucial to analyse the existing data and extract useful information to find the reason for traffic accidents [9]. Mining Techniques is applied to traffic accidents dataset to find valuable information and suggestion on how to drive safely. Data mining results help organizations with transportation, investigate accident data recorded by the police information system, and determine various patterns [10]. The traffic safety issue rose worldwide with significant concerns and has become one of the critical, challenging issues. Traffic accidents have a massive impact on the economy due to the cause of deaths and injuries. Currently, researchers pay more attention to decisive factors that affect significantly traffic accidents. The increase in traffic on the roadway also grows the number of accidents, which is a problem that should be solved, and it is not suited for smart cities. Therefore, it is crucial to analyse the existing data and extract useful information to find the reason for traffic accidents. In purposed research, we analyse the traffic accident dataset to find useful insight and patterns from that data and forecast possible accidents in advance. We use a dataset of the UK Road Accident to find out the casualty severity of accidents. We apply classifications algorithms and association rules on both datasets to find useful information and patterns.

II. LITERATURE REVIEW

Kaur [11] purpose a solution that predicts the cause of the accident, the hilly area, the accident time analysis, the investigated vehicle using explorative visualization techniques. They use R's Integrated Development Environment (R-studio) that is a graphical and statistical calculation tool on accidental road datasets for the analysis. Taamneh [12] use the WEKA tool to evaluate traffic accident data and identify the factors that affect the accident severity. The author applied classifiers PART, DT J48, MLP, and NB to forecast the injury cruelty of traffic accidents based on the Abu Dhabi traffic accident records of 6 years. Tiwari [13] classified traffic accidents according to the category of road users. Pecherková [14] investigate the incident severity according to the different accident circumstances. They use different discrete variables that are vehicle condition, bad structural arrangements, and watchfulness of a driver. Li [15] take five different attributes to find out fatal accidents, they discovered association rules by the Apriori algorithm and found clustering by the K-mean algorithm. They conclude the results that human factors like drunk or not and the type of collision strongly affect the fatal rate. Addi [16] proposed method is distributed into two modules. The first is the extraction module of association rules, using efficient algorithms they can extract rules from the dataset. The second approach is multi-criteria analysis to select the most relevant. Weng [17] proposed my fuzzy association rules to develop an algorithm from unreliable data. Toshniwal [18] K-mean algorithm and the ARM method are applied to solve the traffic accident problem. Different accidental location is divided into three categories: high, medium, and low frequency to extract the hidden information from the dataset and take some preventive action depending on the location of the accident. Solaiman [19] collected accident data in different ways then located it into a centralized database server. Different parameters are used to find the behaviour concerning time [20-21]. They used an API based system, which is more flexible to find the rough and dangerous roads. Krishnaveni [22] applied data mining techniques to perform a prospective analysis of traffic accidents. They focus on some classification models to forecast the severity of injuries in road accidents. Several Mata Classifiers, PART Rule Classifier, Bayesian Naive Bayesian Classifier, J48 Decision Tree Classifier, Forest Tree Random Classifier and AdaBoostM1 were compared to classify the severity of various road accidents. In the end author claim that the Radom Forest works better than the other four algorithms. Tian [23] proposed a data mining technique for the analysis of road traffic accident causes. Nguyen [24] present a state of the art survey which is focused on the classification and clustering through mining algorithms of the data stream.

Zhang [25] extract accident information, reconstruct, and simulate the accident. The proposed method creates the model of the traffic accident scene first and then an accident model is established (cars, block, road, people, etc.) after performing this repeated simulation to ensure that all results are measured in the place of the accident and then based on the actual simulation of the accident, continue to perform simulations changing the conditions, such as vehicles, roads, people, etc., to investigate the road safety of this section of the road. Babic [26] proposed how to use data collected for traffic accidents and find different patterns and factors that are important and cause accidents of various kinds. They use the United Kingdom's Road accident data sample of ten years from 2005 to 2015. Two algorithms were selected: Predictive Mining Decision Tree and Apriori Descriptive Mining Algorithm for generating interesting association rules. In the end, the results achieved are reasonably incomparable to similar work. Kalia [27] present a survey on association rules mining to find out useful patterns from the dataset. Malta [28] conclude that human error is involved in all three-fourths of traffic accidents. They use speech and brake pedal force models to capture the behaviour of the driver better.

Abellan [29] use decision rules to perform the analysis of rural highways of two-lane in Spain and Granada. Kashani [30] analyze crash records using CART and record is collected from the department of information and technology of the Iran traffic police from the years 2006 to 2008. Rovsek [31] analyzed traffic accident data from 2005 to 2009 in Slovenia they use regression tree and classification algorithms. Beshah [32-33] used the basic dataset of Addis Ababa which contains 18,288 accident records. Decision trees, Naïve-Bayes, and k-nearest neighbour's algorithms are used to classify data.

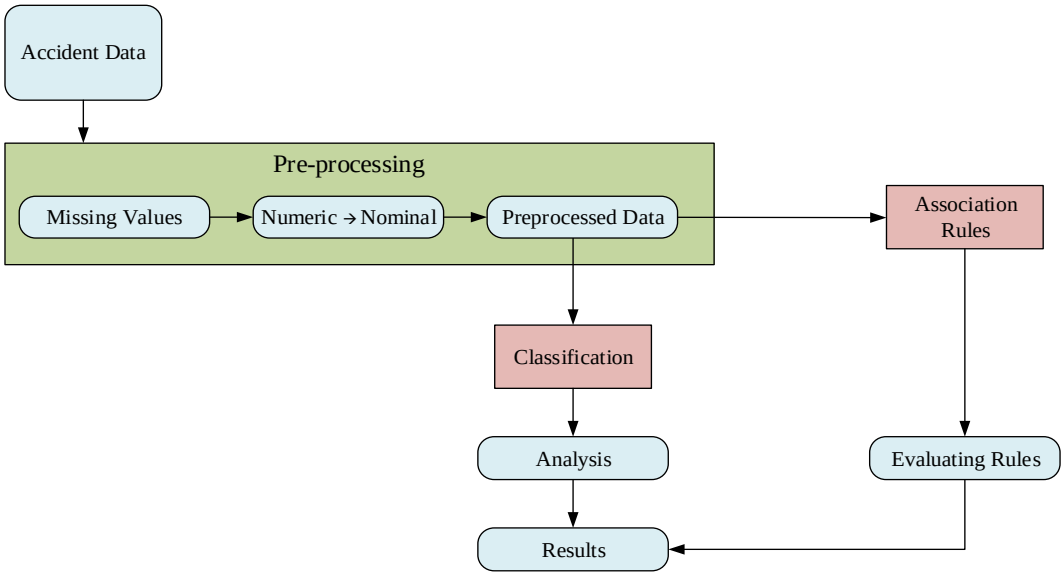

Fig 1. Proposed Model for Road Traffic Accidents Analysis

III. Research Methodology

The steps used in a proposed approach for data analysis are shown in Fig 1. UK Traffic Road Accidents dataset of 2016 is examined in our research which is publicly available and used in different research that is relevant to traffic accidents. A dataset of the year 2016 contains 2549 records and 16 attributes. Data pre-processing is applied because in real-world data is dirty, incomplete, noisy and inconsistent. To achieve the quality mining result in data mining should provide quality data. Attributes that contain missing values in the record were removed, and all numeric values are replaced with nominal values according to the requirement of algorithms. We find the attribute casualty severity of accidents based on three types: Serious, Slight and Fatal.

First, it applied association mining rules to find out relationships between attributes then classification algorithms are applied to find out patterns from the dataset. Weka is used as data analysis tools to perform this analysis.

The result of our analysis contains association rules between the different variables and also provides the patterns that help to reduce the accidents, especially reduce the severity of the accident. The accidents factors used in this research can be summarized in Fig 2.

IV. Statistical Characterization Data Mining

Association rule is a popular technique in data mining that can extract interesting rules from the given dataset. Quality of rule is assessed by various interesting measures. Rule X → Y interesting measures are discussed as follows.

Support (S_P)

The support defines the occurrence of X and Y together, & its percentage is calculated through Eq.1

$$S_p = \frac{P(X \cap Y)}{N} \quad (1)$$

N belong to the total no of accident records.

Confidence (C_f)

The occurrence ratio of X and Y together is confidence, and it is calculated through Eq. (2).

$$C_f = \frac{P(X \cap Y)}{P(X)} \quad (2)$$

Lift (L_t)

Raio of confidence is lifted and excepted confidence of the rule, it can measure the X and Y occurrence together than excepted and calculated through Eq. (3).

$$L_t = \frac{P(X \cap Y)}{P(X) \times P(Y)} \quad (3)$$

Leverage (L_v)
The difference between X and Y appearing together is measures by leverage and calculated using Eq. (4).

$$L_v = P(X \cap Y) - P(X) \times P(Y) \quad (4)$$

Conviction (Cv)
To find out the probability of X if Y occurs or X occur without Y conviction is used and calculated by Eq. (5).

$$C_v = \frac{P(X) \times P(Y)}{P(X \cap Y)} \quad (5)$$

V. Results and Findings

The number of casualty severity is shown in Fig 3. Most accidents happened in November and least in February and the nature of these accidents are slight and most accidents happened in July and October and least in March with slight nature.

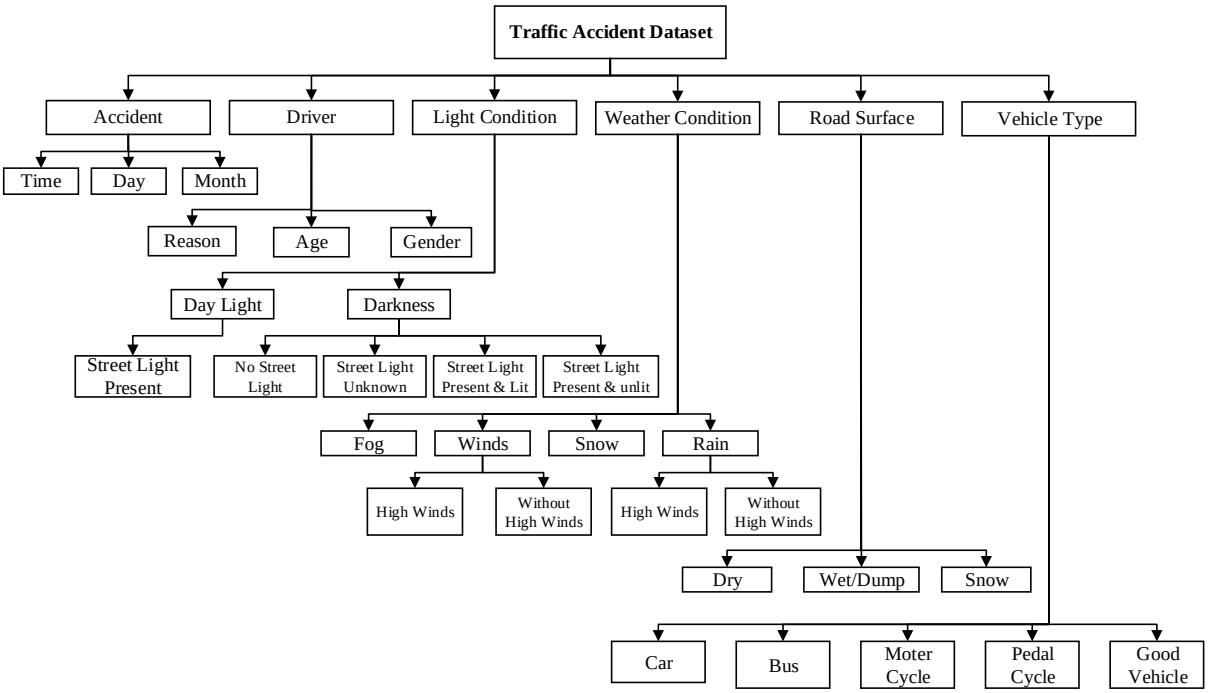

Fig 2. Accident Factors of Road Traffic Accident

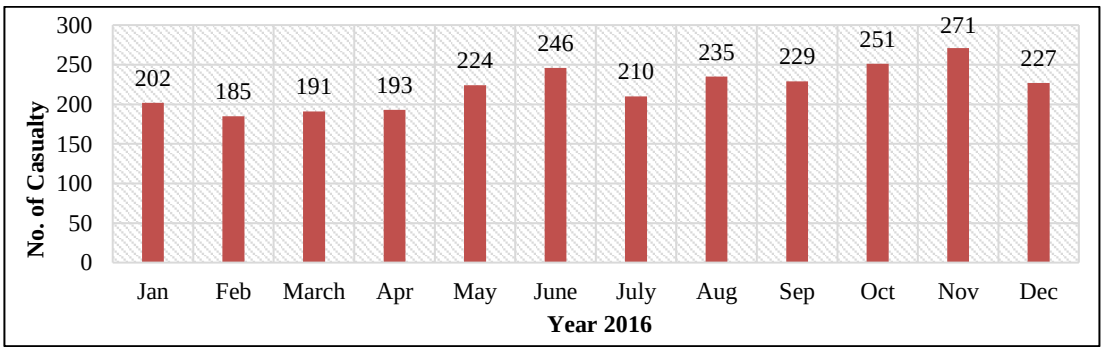

Fig 3. No of Casualty Severity in the Year 2016

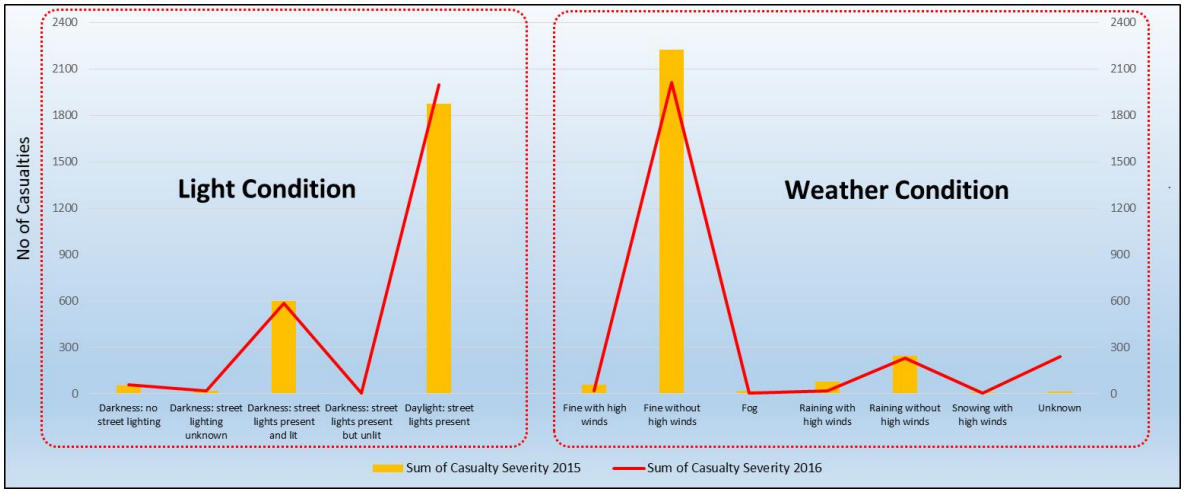

Fig 4. Sum of Casualty Severity in Light and Weather Conditions

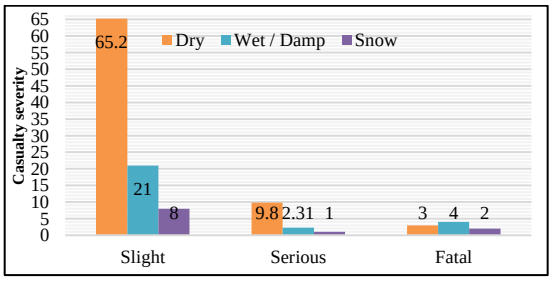

Fig 5. Sum of Casualty Severity in Road Surface

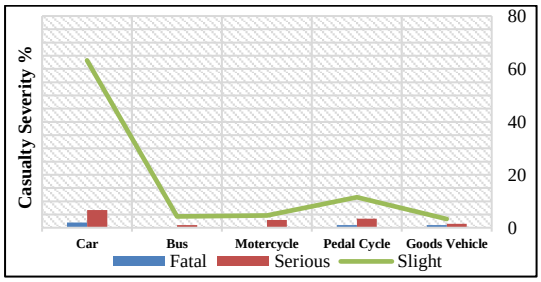

Fig 6. Casualty Severity in Type of Vehicle

Light and Weather Condition:
Fig 4 shows the sum of the casualty by presents the comparison of lighting and weather condition of both years and find out that daylight: the streetlight present is the main cause of casualties. In a dataset of 2015 number of casualties are 2000, but in 2016 it decreases a bit. In the weather conditions of both datasets, the main reason for casualties is fine weather without high. About 2000 casualties happen in both cases.

Roadway Surface Condition:
Fig 5 shows that casualty severity in the road surface, in both cases Dry surface is the cause of slight casualties which is 65%.

Type of Vehicle:
Fig 6 shows the types of vehicles which is involved in casualty severity, in both cases cars are the main cause of slight causalities, it grows 64%.

In Table 1 road traffic accidents attributes of different papers are shown. The light condition, Weather condition and Road Surface these attributes were investigated in almost every paper, but all these attributes cannot investigate in one paper. In the proposed research, almost all attributes which accident date are, time, day, accident severity, collision type, driver attributes, weather condition, light condition, road surface, no of vehicles, type of vehicles are investigated, and results are interpreted through applying classification and association rules. Our empirical results show that the proposed model could classify road traffic accidents with reasonable accuracy.

Association Rule Mining:
The tuples with missing values were removed before applying the algorithms. The numerical values are converted into nominal and clean data is stored in CVS format. Data is analyzed through Weka analysing tool. A clean dataset contains 2664 tuples and 9 attributes. Ten best rules which are shown in Table 2 which are found after applying the Apriori Algorithm in Weka with minimum support of 0.4 and minimum confidence of 0.6.

Table 1: Road Traffic Accidents Attributes of different Papers

Road Traffic Accidents Attributes																	
Sr. No	Authors	High Speed	Accident			Accident Severity	Collision Type	Driver				Seat Belt	Light Condition	Road Surface	No. of Vehicles	Type of Vehicle / Condition	Weather
			Month	Day	Time			Reason	Gender	Age	Experience						
1	Taamneh et al. [4]	×	×	✓	✓	×	×	✓	✓	✓	✓	✓	✓	✓	×	×	×
2	Tiwari et al. [5]	×	✓	✓	✓	×	×	×	×	✓	×	×	✓	✓	✓	✓	✓
3	Pecherková et al. [6]	×	×	×	×	×	×	✓	✓	✓	✓	✓	×	✓	×	✓	×
4	Li et al. [7]	✓	×	×	×	×	✓	×	×	×	×	×	✓	✓	×	×	✓
5	Addi et al. [8]	×	×	×	×	×	✓	✓	×	✓	✓	×	✓	×	×	✓	✓
6	Kumar et al. [9]	×	✓	✓	✓	✓	✓	×	×	×	×	×	✓	✓	×	×	×
7	Krishnaveni et al. [11]	×	×	×	×	✓	×	×	×	×	×	×	×	✓	✓	×	×
8	Tian et al. [12]	×	×	×	×	×	×	✓	✓	✓	✓	✓	✓	✓	×	✓	✓
9	Chang et al. [13]	✓	×	✓	×	×	×	×	×	×	×	×	✓	✓	✓	×	✓
10	Zhang et al. [14]	✓		✓									✓	✓	✓		✓
11	Beshah et al. [20]	×	×	×	×	✓	×	×	×	×	×	×	✓	✓	×	×	✓
12	Proposed Research	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓

Table 2: Association rules discovered by Apriori Algorithm

Sr. No	Rules	Confidence
1	**IF** { CC = DR **AND** TOV = C }sss **THEN** { CS = SL }	0.94
2	**IF** { SOC = F **AND** TOV = C } **THEN** { CS = SL }	0.94
3	**IF** { LC = DL (SLP) **AND** SOC = F } **THEN** { CS = SL }	0.92
4	**IF** { WC = Fw-oHW **AND** SOC = F } **THEN** { CS = SL }	0.92
5	**IF** { RS = D **AND** LC = DL (SLP) **AND** TOV = C } **THEN** { CS = SL }	0.92
6	**IF** { RS = D **AND** LC = DL (SLP) **AND** WC = Fw-oHW **AND** TOV = C } **THEN** { CS = SL }	0.92
7	**IF** { LC = DL (SLP) **AND** WC = Fw-oHW **AND** TOV = C } **THEN** { CS = SL }	0.91
8	**IF** { LC = DL (SLP) **AND** TOV = C } **THEN** { CS = SL }	0.91

Association rules find out the severity of casualty with confidence. Where CC = Casualty Class, DR = Driver / Rider, RS = Road Surface, D = Dry, WC = Weather Conditions, Fw-oHW = Fine without high winds, CS = Casualty Severity, SL = Slight, LC = Light Conditions, DL (SLP) = Daylight: Street lights present, TOV = Type of Vehicle, SOC = Sex of Causality, F = Female and C = Car.

Classification

In classification, the ZeroR classifier is used with 10 cross-validation folds, and it gives us an 87% accuracy rate.

The overall prediction accuracy using different techniques is presented in Fig 7. ZeroR shows that the severity of causality is mostly slight. We also apply association rules and classification on both datasets to find out the severity of casualty and find some rules using Apriori Algorithm then apply the ZeroR classifier to find accuracy rate and some other measures.

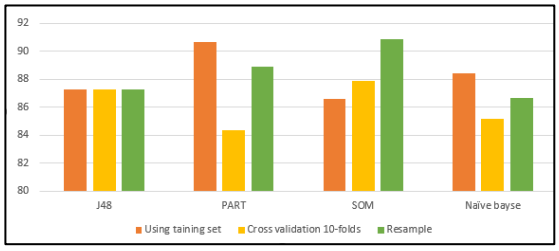

Fig 7. Overall Prediction Accuracy Using Different Techniques

VI. CONCLUSIONS

In this study, we find casualty severity from the data set based on different attributes like road surface, weather condition, light condition, and type of vehicles. We have also applied classification on the dataset to predict the severity of casualty with reasonable accuracy. Some Association rules were discovered by the using Apriori algorithm that could

be used by the analyst to identify the main reasons that contribute to the severity of causality. Four classifiers were built, which are Naïve Bayes, decision tree, support vector machine, and rule induction and obtained accuracy based on using training data, cross-validation 10-folds, and resampling. The most critical factors associated with a person in severe accidents are road surface, light and weather condition, gender, and type of vehicle. In the future, we will improve the study with the help of a large dataset.

REFERENCES

[1] Grossi, V., A. Romei, and F. Turini, *Survey on using constraints in data mining.* Data Mining and Knowledge Discovery, 2017. **31**(2): p. 424-464.

[2] WHO. *World Health Organization*. 2017 30 October, 2017]; Available from: http://www.who.int/gho/road_safety/en/.

[3] Perez, K., et al., *Implications of estimating road traffic serious injuries from hospital data.* Accident Analysis & Prevention, 2019. **130**: p. 125-135.

[4] Naqvi, N.K., M.A. Quddus, and M.P. Enoch, *Do higher fuel prices help reduce road traffic accidents?* Accident Analysis & Prevention, 2020. **135**: p. 105353.

[5] Deebak, B.D. and F. Al-Turjman, *A hybrid secure routing and monitoring mechanism in IoT-based wireless sensor networks.* Ad Hoc Networks, 2020. **97**: p. 102022.

[6] Blog, M.I.L. *Accident Statistics*. 2017 30 October, 2017]; Available from: https://www.marylandinjurylawyerblog.com/car_accident_statistics_from_t.html.

[7] Williams, P.K., *A Clustering Rule Based Approach for Classification Problems.* International Journal of Data Warehousing and Mining, 2012. **1**(8): p. 23.

[8] Al-Turjman, F., *Information-centric framework for the Internet of Things (IoT): Traffic modeling & optimization.* Future Generation Computer Systems, 2018. **80**: p. 63-75.

[9] Al-Turjman, F. and C. Altrjman, *Enhanced Medium Access for Traffic Management in Smart-cities' Vehicular-Cloud.* IEEE Intelligent Transportation Systems Magazine, 2020.

[10] Sahoo, J., A.K. Das, and A. Goswami, *An effective association rule mining scheme using a new generic basis.* Knowledge and Information Systems, 2015. **43**(1): p. 127-156.

[11] Kaur, G. and H. Kaur, *Black Spot and Accidental Attributes Identification on State Highways and Ordinary District Roads Using Data Mining Techniques.* International Journal, 2017. **8**(5).

[12] Taamneh, M., S. Alkheder, and S. Taamneh, *Data-mining techniques for traffic accident modeling and prediction in the United Arab Emirates.* Journal of Transportation Safety & Security, 2017. **9**(2): p. 146-166.

[13] Tiwari, P., S. Kumar, and D. Kalitin. *Road-User Specific Analysis of Traffic Accident Using Data Mining Techniques*. in *International Conference on Computational Intelligence, Communications, and Business Analytics*. 2017. Springer.

[14] Pecherková, P. and I. Nagy. *Analysis of discrete data from traffic accidents*. in *Smart City Symposium Prague (SCSP), 2017*. 2017. IEEE.

[15] Li, L., S. Shrestha, and G. Hu. *Analysis of road traffic fatal accidents using data mining techniques*. in *Software Engineering Research, Management and Applications (SERA), 2017 IEEE 15th International Conference on*. 2017. IEEE.

[16] Al-Maweri, N.A.A.S., et al., *State-of-the-art in techniques of text digital watermarking: challenges and limitations.* Journal of Computer Science, 2016. **12**(2): p. 62-80.

[17] Weng, C.-H. and Y.-L. Chen, *Mining fuzzy association rules from uncertain data.* Knowledge and Information Systems, 2010. **23**(2): p. 129-152.

[18] Kumar, S. and D. Toshniwal, *A data mining approach to characterize road accident locations.* Journal of Modern Transportation, 2016. **24**(1): p. 62-72.

[19] Solaiman, K., M.M. Rahman, and N. Shahriar. *AVRA Bangladesh collection, analysis & visualization of road accident data in Bangladesh*. in *Informatics, Electronics & Vision (ICIEV), 2013 International Conference on*. 2013. IEEE.

[20] Kavitha, S., et al., *An Internet Of Things For Data Security In Cloud Using Artificial Intelligence.* International Journal of Grid and Distributed Computing, 2021. **14**(1): p. 1257-1275.

[21] Lv, Z., et al., *AI-empowered IoT security for smart cities.* ACM Transactions on Internet Technology, 2021. **21**(4): p. 1-21.

[22] Krishnaveni, S. and M. Hemalatha, *A perspective analysis of traffic accident using data mining techniques.* International Journal of Computer Applications, 2011. **23**(7): p. 40-48.

[23] Tian, R., Z. Yang, and M. Zhang. *Method of road traffic accidents causes analysis based on data mining*. in *Computational Intelligence and Software Engineering (CiSE), 2010 International Conference on*. 2010. IEEE.

[24] Nguyen, H.-L., Y.-K. Woon, and W.-K. Ng, *A survey on data stream clustering and classification.* Knowledge and information systems, 2015. **45**(3): p. 535-569.

[25] Zhang, Y. and T. Zou. *An Analysis of Road Segment Traffic Safety Based on Accident Reconstruction*. in *Measuring Technology and Mechatronics Automation (ICMTMA), 2016 Eighth International Conference on*. 2016. IEEE.

[26] Babič, F. and K. Zuskáčová. *Descriptive and predictive mining on road accidents data*. in *Applied Machine Intelligence and Informatics (SAMI), 2016 IEEE 14th International Symposium on*. 2016. IEEE.

[27] Kalia, H., *A Survey on Fuzzy Association Rule Mining.* International Journal of Data Warehousing and Mining, 2013. **1**(9): p. 27.

[28] Malta, L., C. Miyajima, and K. Takeda, *A study of driver behavior under potential threats in vehicle traffic.* IEEE Transactions on Intelligent Transportation Systems, 2009. **10**(2): p. 201-210.

[29] Abellán, J., G. López, and J. De OñA, *Analysis of traffic accident severity using decision rules via decision trees.* Expert Systems with Applications, 2013. **40**(15): p. 6047-6054.

[30] Tavakoli Kashani, A., A. Shariat-Mohaymany, and A. Ranjbari, *A data mining approach to identify key factors of traffic injury severity.* PROMET-Traffic&Transportation, 2011. **23**(1): p. 11-17.

[31] Rovšek, V., M. Batista, and B. Bogunović, *Identifying the key risk factors of traffic accident injury severity on Slovenian roads using a non-parametric classification tree.* Transport, 2017. **32**(3): p. 272-281.

[32] Beshah, T. and S. Hill. *Mining Road Traffic Accident Data to Improve Safety: Role of Road-Related Factors on Accident Severity in Ethiopia*. in *AAAI Spring Symposium: Artificial Intelligence for Development*. 2010.

[33] Beshir, A.A. and J. Song, *Urbanization and its impact on flood hazard: the case of Addis Ababa, Ethiopia.* Natural Hazards, 2021: p. 1-24.