PREDICTING TRAFFIC ACCIDENTS AND THEIR INJURY SEVERITIES USING MACHINE LEARNING TECHNIQUES

H. L. GURURAJ[1], V. JANHAVI[2], U. TANUJA[2], FRANSESCO FLAMINI[3], B. C. SOUNDARYA[2], VINAYAKUMAR RAVI[4]
[1]Department of Information Technology, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, India
[2]Vidyavardhaka College of Engineering, India
[3]IDSIA USI-SUPSI, University of Applied Sciences and Arts of Southern Switzerland, Switzerland
[4]Center for Artificial Intelligence, Prince Mohammad Bin Fahd University, Saudi Arabia

ABSTRACT
Traffic accidents are among the most censorious issues confronting the world as they cause numerous deaths, wounds and fatalities just as monetary misfortunes consistently. According to the world health organization (WHO) reports, 5,18,3626 accidents took place in India in the year 2019. Factors that contribute to these road crashes/ traffic accidents and resulting injuries include inattentive drivers, unenforced traffic laws, poor road infrastructure, driving in bad weather conditions and others. This investigation effort establishes models to select a set of influential factors and to build up a model for classifying the severity of injuries. Machine learning models can be applied to model and predict the severity of injury that occurs during road accidents. One such way is to apply unsupervised learning models such as Apriori, Apriori TID (transaction id), SFIT (set operation for frequent itemset using transaction database) and ECLAT (equivalence class clustering and bottom-up lattice traversal) which analyze the unlabeled traffic accidents dataset and determine the relationship between traffic accidents and injury. This research work is helpful for traffic departments to decrease the number of accidents and to distinguish the injury's seriousness extensive simulations were carried out to demonstrate the unsupervised learning algorithms for predicting the injury severity of traffic accidents. Apriori algorithm predicts the patterns in 962 milliseconds, Apriori TID (transaction id) algorithm predicts the pattern in 557 milliseconds, SFIT algorithm predicts the pattern in 516 milliseconds and ECLAT algorithm predicts the pattern in 124 milliseconds. ECLAT algorithm took less time compared to all the other algorithms.
Keywords: Apriori, ECLAT, machine learning, traffic accidents

1 INTRODUCTION

Traffic accidents cause considerable economic, monetary and social losses to individuals, families and the nation as a whole. As indicated by the death approximation provided by the world health organization (WHO), there is an alarming increase in the number of traffic accidents [1][2]. Each year the lives of approximately 1.2 million are lost with 50 million individuals injured [1].

Road traffic accidents have reached alarming levels every year across the world. Every year, roughly 1.2 million people are dies due to road traffic crashes worldwide. A similar trend can be seen in countries with the most traffic accident-related death rate per 100,000 namely Zimbambwe (61.90), Liberia (52.03), Malawi (51.62), Gambia (47.51), Togo (46.62), Tanzania (46.17), Rwanda (45.90), Sao tome (45.52), Burkina Faso (44.94) and Burundi (44.94) [3]. The public is suffering from many major injuries even after many years of the accident [4]. Consequently, road crashes became the main source of human death and injuries all around the world.

There are several factors due to which accidents occur. These are: rear-end collisions, side-impact collisions, side swap collisions, rollover, hit and run, head-on collisions, single-car accidents, multiple vehicle pile-ups and drunk and drive. There are two types of accidents namely: Major accident injury and minor accident injury. Major accident injury is one in

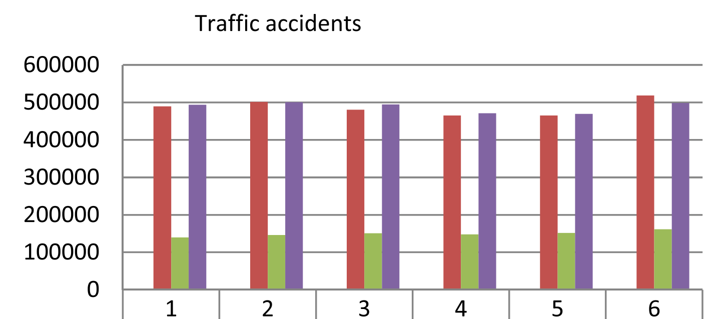
ISSN: 2058-8305 (paper format), ISSN: 2058-8313 (online), http://www.witpress.com/journals
DOI: 10.2495/TDI-V6-N4-363-377

which at least one casualty endures genuine injury requiring hospitalization. Minor accident injury is one in which the victim(s) doesn't require hospitalization.

Figure 1 represents the graph of traffic accidents categorized by three parameters i.e., road accidents, the person dies and injuries that happened on roadways from 2014 to 2019, the data provided by the Government of India (ministry of road transport and highways research wing, New Delhi). The descriptive model was developed using clustering techniques and association learning techniques [5][6]. Around 140 lives have been lost across the country due to road crashes. Different states which recorded maximum traffic accidents are Delhi, Maharashtra, Gujarat, Assam, Kerala, Karnataka, Rajasthan, Punjab and Tamil Nadu [9][10].

Figure 2 represents the different traffic accidents with related injuries. For example, being drunk and driving may lead to some fractures or may lead to death. Different types of injuries are brain injuries, spine fracture, pelvic fracture, back and spinal cord trauma, skull and maxillofacial, rib fracture, broken bones, whiplash, scrapes and cuts, internal bleeding, and herniated disc, knee trauma, soft tissue injuries, etc. The existing system used some of the machine learning algorithms like the random forest (RF), K-nearest neighbor and decision tree to identify the accidents, accidents vehicles, and injuries which provides less accuracy [7][8].

The proposed work is a real-time application that is helpful for the government sector to decrease the number of traffic accidents and its injuries severity. Unsupervised learning is a machine learning technique, which helps to discover the hidden patterns and information in the given datasets. Mainly unsupervised learning algorithm needs only data sets without trained or labeled and tested which does not consume time for calculating, analyzing and predicting accidents severities. We use association rule mining which predicts the severities of road crashes [11][14]. Association rule mining is used to discover the relationships between data items within large datasets. Association rule uses support and confidence to identify the most important relationships of data items in less amount of time. The main idea of this work



	2014	2015	2016	2017	2018	2019
Year	2014	2015	2016	2017	2018	2019
Road accidents	4,89,40	5,01,42	4,80,65	4,64,91	4,64,91	5,18,32
Persons Death	1,39,67	1,46,13	1,50,78	1,47,91	1,51,41	1,61,51
Persons injured	4,93,47	5,00,27	4,94,62	4,70,97	4,69,41	4,98,40

Figure 1: Trends of traffic accidents, the person dies, and injuries by road category from 2014 to 2019.

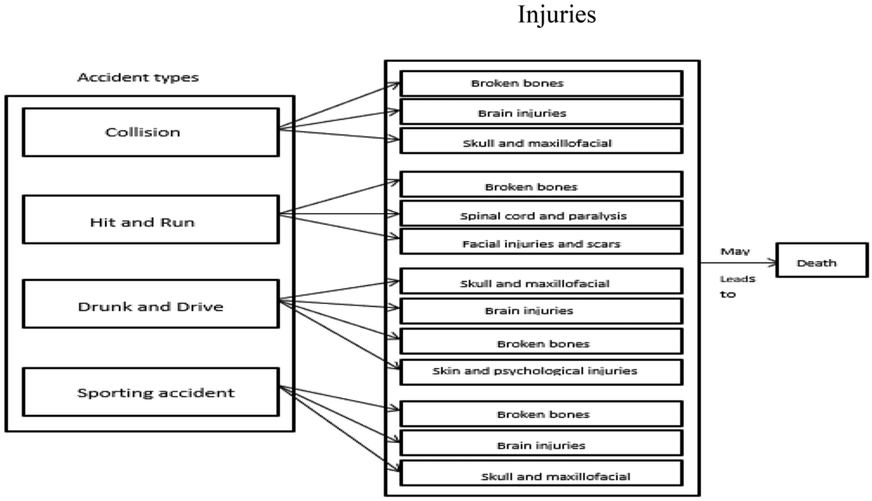


Figure 2: Different traffic accidents with injuries.

is to minimize traffic accidents and their injuries because the public is suffering from many major injuries even after many years of accidents [17][18]. We took four machine learning algorithms i.e., Apriori, Apriori TID (transaction id), SFIT (set operation for frequent itemset using transaction database) and ECLAT (equivalence class clustering and bottom-up lattice traversal) which analyze the unlabeled traffic accidents dataset and determine the relationship between traffic accidents and injury.

This finding of the paper is as follows:

- Identify the correlation between the accident types i.e., drunk and drive may lead to hit and run, etc.
- Discovering the relationship between the different categories of injuries i.e., brain injury may lead to spinal cord injury, etc.
- Identifying the correlation between accident types with injuries i.e., collision may lead to brain injury, broken bones, etc.

Our work can be helpful in several future scenarios to save people's lives by early prediction of the type of accidents with related injuries to get proper treatment because so many people are suffering from major injuries even after many years of the accident. The paper is organized as follows: section 2 represents related work, section 3 describes proposed work, section 4 presents a comparative analysis of existing methodologies, section 5 shows the results and section 6 presents a comparative analysis of proposed methodologies. Finally, we concluded this paper.

2 RELATED WORK

In this section, we briefly discussed existing methodologies and their contributions. We also acknowledge the limitations of existing methodologies.

Xiao Li, et.al, [11] identify the impact of traffic accidents and injury severities which is dependent on three angles, example, daily travelers (included age, sex, and so forth.), vehicles (included vehicle type and number, vehicle transmission and so on.) and street (included pavement condition, cross-area type and so forth.).

Siddharth Tripathi, et.al, [12] proposed a model called CBIT (cloud-based intelligent traffic system) to detect traffic accidents and alerts the traffic authorities of the location of an accident. They took three principle targets: examining emission, accident detection and unique vehicle ID. The proposed model used some sensors i.e., NO_X, SO_X, CO and temperature sensors like MQ135, MQ3 and LM35 for detecting the emission. Accelerometer (ADC MCP3008) and force resistive sensors are used to detect accidents. Unique vehicle ID will be assigned to all the vehicles which are helpful for traffic authorities to monitor the vehicle. GPS (UBLOX G7020) and GSM (SIM 900) modules are used for sharing the location of accidents with traffic authorities. The WiFi port of raspberry pi is used to aggregate the information and then it is stored in the cloud.

Naji Taaib Said Al Wadhahi, et.al, [13] proposed a model for the detection of accidents and prevention system to lessen traffic hazards utilizing Arduino boards (Atemga 328) and some sensors such as IR. This framework includes two stages: detection and prevention of traffic accidents. Here is a sensor (IR) that helps to identify the accidents and alarm the individuals by sending messages. Messages are sent through GSM (SIM 900D) module. These sensors notify the driver about accidents.

JIAN ZHANG, et.al, [14] compared injuries severities predictions using statistical methods and machine learning algorithms. The accident seriousness, street geometry and traffic status were gathered at road separate territories in Florida [24]. They assessed the two most utilized statistical methods which are the multinomial logit model and the ordered probit (OP) model. They have used machine learning algorithms i.e., K nearest neighbor (KNN), decision tree, RF and support vector machine (SVM) to predict the accident injury severities. The RF method is best for large and extreme accidents while the OP model was the most fragile one.

Guang Yu, et.al, [15] proposed a strategy for taking gander blockage and controlling traffic accidents. Speed and lane changing control systems are used to reduce traffic accidents. They took a portion of the countermeasures to keep away from the auto collisions through administration of path, controlling rate of crashes, adjusting of speed and traffic data.

Fabio Galatioto, et.al, [16] proposed a model i.e., MAIA (model and methods for accident prediction and its impact assessments) toolkit for predicting traffic accidents and their injury severities. The authors in this paper concentrated on three different problems: accident rate prediction, estimation of both injury and non-injury collisions and estimation of collision security level. Finally, we discussed existing works and their contributions. We surveyed so many papers and their solutions which are related to our work. There are so many issues in existing work, so we compare these solutions with our proposed solutions. Our work has made more contributions to these issues.

3 PROPOSED WORK

Road traffic accidents and their injuries are the leading cause of death. The proposed research work is a real-time application that is helpful for traffic departments to decrease the number of traffic accidents and distinguishes the injury severities. This application is also helpful for doctors to give better treatment. The proposed system makes use of association learning methods to discover the hidden patterns between traffic accidents and related injury severities.

Figure 3 shows the design for anticipating traffic injury severities. The system predicts the traffic accidents and their injury severities dependent based on old datasets. The users here will be the Admin, Member (City Traffic In Charger, Doctors) and Visitor (Public). The administrator maintains the entire system. The administrator is responsible for the city in charge of creating and uploading the necessary data for processing. Member (City Traffic In

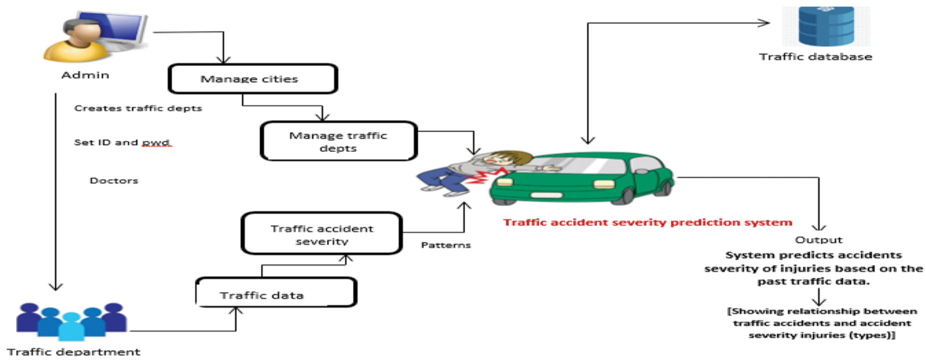


Figure 3: Architecture diagram of traffic accident severity.

Charger, Doctors) City In Charge is a service receiver and responsible for uploading the traffic accidents data into the server. Visitor (Public) Visitor is a user who visits the application. The visitor has limited accessibility.

3.1 Methodology

Unsupervised learning methods help to extract the relevant information and structure in unlabeled datasets. It takes less time to extract the patterns from the given datasets. Association rule discovers the patterns using support and confidence [19]. In this paper, we are using four types of algorithms i.e., Apriori [20], Apriori TID (transaction id) [21], SFIT (set operation for frequent itemset using transaction database) [22,29] and ECLAT (equivalence class clustering and bottom-up lattice traversal) [26] to discover the patterns of traffic accidents and its injury severities.

3.1.1 Apriori algorithm

Apriori algorithm is one of the association rule learning [20]. Defining association rule mining is as follows:

Let $I_{item} = \{i_1, i_2, i_3, \ldots\ldots, i_n\}$ set of n attributes (items). Let $D_{transactions} = \{t_1, t_2, t_3, \ldots\ldots, t_n\}$ set of n transactions (database). Different attributes are used in the Apriori algorithm i.e., Itemset, K-itemset and null sets where the collection of more items including zero is called itemsets. Itemsets include k-items is called k-itemsets. Itemset that does not include any items called null (empty) set.

In each transaction *T* contains a subset of items taken from *I*. Transaction *T* includes an item set *A*. If *A* is a subset of *T* then the support count σ (*A*) can be written mathematically,

$$\sigma(A) = |\{t_i | A \subseteq t_i, t_i \in T\}| \quad (1)$$

Where | | indicated the number of items in a set. Transactions of items in an itemset with support [32] [33].

$$S(A) = \sigma(A) / N$$

Where S(A) is greater than or equal to some defined threshold.

An association rule for the Apriori algorithm with an expression form A->B where A and B are disjoint itemsets i.e., $A \cap B = \emptyset$. It can be measured by using support and confidence. The definition of support and confidence are:

$$\text{Support s(A -> B)} = \frac{\sigma(A \cup B)}{N} \tag{2}$$

$$\text{Confidence c(A -> B)} = \frac{\sigma(A \cup B)}{\sigma(A)} \tag{3}$$

In the above equation (i) and (ii) where support helps to measure the rule in a given itemset and confidence discovers frequently appearing items in a transaction T.

Figure 4 shows the processing of the Apriori algorithm. The data processing means extracting relevant data from the accident database or server. Now applying the Apriori algorithm for generating rules to predict the accident patterns and represents the results of generated patterns.

Algorithm 3.1 Apriori Algorithm for generating frequent itemsets

STEP 1: Examine old data sets then regulate the support (s) for each item.
STEP 2: Create L_1 (only one item set).
STEP 3: First use L_{k-1}, and then unite L_{k-1} to create candidate K sets.
STEP 4: Examine the generated candidate K item set and then create support for the candidate in each item set K.
STEP 5: Append item set frequently, till C = Null set (ø).
STEP 6: Create all non–empty subsets for frequent itemsets.
STEP 7: Examine the confidence for all non-empty subsets. If the examined confidence is higher or equal to the given confidence, then append to the association rule.

3.1.2 Apriori TID Algorithm

Apriori algorithm discovers all the patterns or relations in the given item set. It also discovers the number of candidate items [21].

Suppose $T_{transactions}$ $\{t_1, t_2, t_3, \ldots t_k\}$, $(K \geq 1)$ (transaction sets, T_{item} $(T_i) = (I_1, I_2, I_3, \ldots, I_m)$, $(m \geq 1)$ itemsets and K-itemsets $(I_n) = \{i_1, i_2, i_3, \ldots i_n\}$, $(n \geq 1)$ k-itemsets where $K_{itemsets} \subseteq I$.

Figure 5 shows the processing of an Apriori TID (transaction id) algorithm. Examine the old data items to regulate the support(s) for each item. Find C1, C2, C3,…..Ck and then

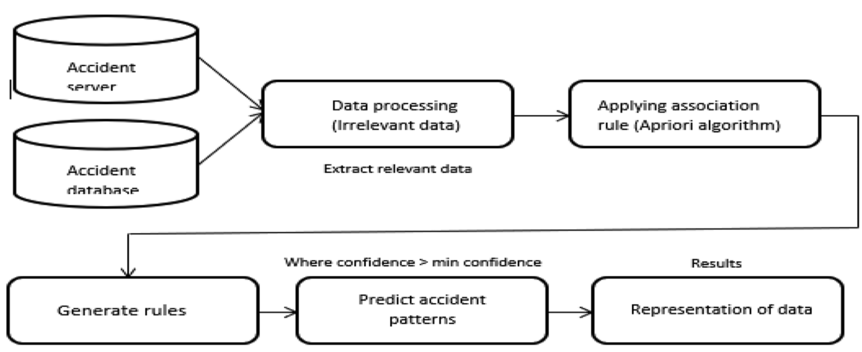


Figure 4: Processing of Apriori algorithm.

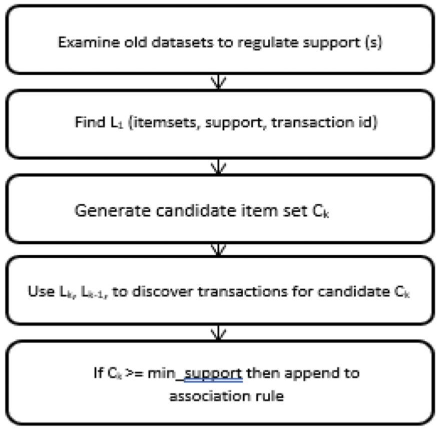


Figure 5: Processing of Apriori TID (transaction id) algorithm.

generate L_1, L_2, L_3,…..L_{k-1} for frequent item set. Generate candidate C_k set by using L_{k-1}. Append item set frequently till C = Null set (ø). If C_k ≥ min-support then append to the association rule and displays the results of generated patterns [30][31].

Algorithm 3.2 Apriori TID Algorithm for generating frequent itemsets

STEP 1: Examine old data sets then regulate the support (s) for each item.
STEP 2: Find C1'
STEP 3: Generate L1 (Frequent item set).
STEP 4: Find C2', C3' …..
STEP 5: First use L_{k-1}, and then unite L_{k-1} to create candidate K sets.
STEP 6: Examine the generated candidate K item set and then create support for the candidate in each item set K, by comparing with the previous step (but not with the original dataset as we did in the Apriori algorithm).
STEP 7: Append item set frequently, till C = Null set (ø).
STEP 8: Create all non–empty subsets for frequent itemsets.
STEP 9: Examine the confidence for all non-empty subsets. If the examined confidence is higher or equal to the given confidence, then append to the association rule.

3.1.3 SFIT (Set operation for frequent itemset using transaction database)

SFIT algorithm is a combination of apriori algorithm which is used for data mining and different set operations like union and intersection are used [22].

For constructing K-itemsets, we use frequent itemsets (K-1). Union is formed for K and (K-1) itemsets. Employing intersection operation for transaction identifiers (tids) of itemsets.

Itemsets {M} transactions with Tid 2, 4, 6, 8 and {N} transactions with Tid 1, 2, 3, 5, 6, 7, 8 i.e., T(M) = {2, 4, 6, 8} and T(N) = {1, 2, 3, 4, 5, 6, 7, 8}. The itemsets {M, N} is the union of MN itemsets. In order to find out the tids for {M, N} by using intersection principle as shown below:

$$T(MN) = T(M) \cap T(N)$$

$$= \{2, 4, 6, 8\} \cap \{1, 2, 3, 4, 5, 6, 7, 8\}$$

$$T(MN) = \{2, 4, 6, 8\}$$

If the obtained output is greater than the given minimum support, then it will be included in the frequent itemset otherwise it will be eliminated [29].

Algorithm 3.3 SFIT Algorithm for generating frequent itemsets

STEP 1: Examine old data sets then regulate the support (s) for each item.
STEP 2: Find C1 (candidate item).
STEP 3: Generate L1 (based on C1 finding L1 i.e., frequent one item set).
STEP 4: Find C2, C3…….., (based on C2……..Cn finding L2…..Ln).
STEP 5: Find C4, when C4 = ø (null set) then stop the iteration.
STEP 6: Find frequent itemset (L) and subsets showing the relationship between items.
STEP 7: Finally compare with minimum specified confidence, the rules where confidence is greater than or equal, add to the strong association rule.

3.1.4 ECLAT (Equivalence class clustering and bottom-up lattice traversal)

ECLAT algorithm is an acronym for equivalence class clustering and bottom-up lattice traversal [24][25]. In ECLAT algorithm, it will make use of vertical transactions id (tid) sets inside the database, clustering of equivalence classes and lattice traversal (bottom-up approach). This algorithm reduces storage and cost [26]. ECLAT algorithm metamorphose the parallel database into perpendicular database i.e., from item set format Apriori<TID_x, X_1, X_2,…..,X_k> to transaction id(tid) set format ECLAT<X_k, TID_1, TID_2,……,TID_k>. ECLAT algorithm uses a vertical approach tidset database includes a list of items: the set of all transaction identifiers called tidset of A, which is represented as tidset(A) = $\{T_i, T_{id} \mid T_i \in Di\ X \subseteq T_i\}$. Several elements in tidset (A) are called support of A i.e., represented as σ (A) = |tidset (A)|.

Figure 6 shows the processing of an ECLAT algorithm. From the accident database or server extracting the relevant information from the database is called data processing then applying the éclat algorithm for generating association rules that predict the accident patterns and represent the results of generated accident patterns.

Algorithm 3.4 ECLAT Algorithm for generating frequent itemsets

Step 1: Generate a tidlist for every item by scanning the database.
Step 2: Tidlist of {a} is precisely the item set of transactions {a}.
Step 3: Bisect the tidlist of {a} with the other items of tidlist, the resulting of dividing tidlist is {a,b}, {a,d}, {a,c}…..

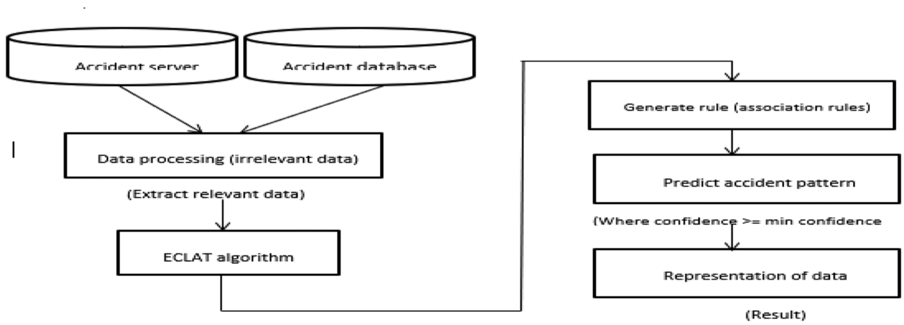


Figure 6: Processing of ECLAT algorithm.

Step 4: Recurrent 1 on {a} – given in the database.
Step 5: Replicate these steps for all other item sets.

4 COMPARATIVE ANALYSIS OF THE EXISTING METHODOLOGIES

In this section, Table 1 shows the performance of existing methodologies.

Table 1: Comparative analysis of other methodologies.

No.	Data sets used	Algorithms or techniques used	Accuracy	Limitations
[11]	>300 datasets of traffic accidents.	KNN, decision tree and association rule.	KNN – 80.26% Decision tree – 73.68%	Small Dataset, hence, less accurate results.
[12]	500 traffic accident data.	Raspberry pi, emission sensors, GPS and GSM modules.	70.1%	The cost of hardware components is very high.
[13]	5,000 datasets of traffic accidents	KNN, decision tree, RF, SVM.	Overall accuracy ranges from 44.7% to 80.5%	Used labeled/ trained datasets that lead to very low accuracy.
[14]	>6,000 datasets of Michigan traffic accidents.	RF, logistic regression, naïve bayes and ada boost.	Logistic regression – 74.5% Naïve Bayes – 73% Ada boost – 74.5% RF – 75%	This model is using only Michigan datasets and has less accuracy.
[15]	3,643 traffic accident data of china.	Bayesian network and information entropy.	Overall accuracy ranges from 50 to 90%	Using fewer china datasets leads to less accuracy.
[16]	1,130 traffic accident datasets.	Convolution neural network (CNN).	87%	Used fewer amounts of datasets which leads to low accuracy.
[17]	7,000 traffic crashes datasets of the UK	Feedforward neural network (FNN), SVM, fuzzy c means clustering-based feed-forward neural network (FNN-FCM) and fuzzy c means based support vector machine (SVM-FCM).	FNN – 69% FNN-FCM – 70.50% SVM – 73% SVM-FCM – 74%	They used both training and testing datasets but the accuracy is less compared to the proposed methodology.

Table 1: (*Continued*)

No.	Data sets used	Algorithms or techniques used	Accuracy	Limitations
[18]	150 traffic accident trained datasets.	Convolution neural network (CNN).	78.5%	Even though the datasets are small, this method gives less accuracy.
[19]	>10,000 traffic accident datasets of USA.	Naïve bayes, RF, MLP, Ada boost.	NB – 74% RF – 77% MLP – 77% AB – 75%	This framework predicts traffic accident severity in the USA but the accuracy is less.
[20]	> 4,000 traffic accident datasets.	Ordered a probit model, an artificial neural network (ANN) and an ensemble model.	OP – 80% ANN – 86.50% EM – 87%	Accuracy is less compared to the proposed methodology.

5 RESULTS ANALYSIS

In this paper, all the below figures represents the outcome of our proposed methodology. For predicting the injury patterns of traffic accidents, we are using unsupervised learning or association mining rules. The unsupervised learning method gives relevant information from the given datasets without training and testing those datasets. We took four algorithms to predict the traffic accident severities. All the four algorithms Apriori, Apriori TID, SFIT and ECLAT generate the patterns in three ways i.e., firstly, accident type with injuries for example collision may lead to spine fracture with 75% confidence means that 75% possibility of people frequently met with this accident severity. Second, discovering the relationship between the different categories of injuries for example brain injury may lead to spinal cord injury with 85% of confidence. The third category is to find out the relation between the accident types for example drunk and driving may lead to hit and run with 90% of confidence. Almost all the algorithms will generate the same kind of patterns but the difference is efficiency will vary. In this paper, we find out that the éclat algorithm is the best mining method to generate the patterns because it takes less storage capacity, cost and time.

Figure 7 shows the prediction of traffic accident severity patterns using ECLAT algorithm. This algorithm uses bottom up approach for mining frequent itemsets. ECLAT algorithm does not scan whole database to generate support values. The output of this ECLAT algorithm showed in three ways i.e., accident type with injuries (collision may leads to brain injuries and broken bones with confidence 100%), discovering the relation between different type of injuries (brain injury may leads to broken bones with confidence 100%) and also it helps to find out the accident types (drunk and drive may leads to hit and run with 100% confidence). It is best suitable for large and small datasets and discovers the pattern in 124 milliseconds as shown in above figure. This algorithm is best for mining when compare to Apriori, Apriori TID and SFIT algorithm because it takes less storage capacity, cost and less time.

As per the survey of the World health organization (WHO), Figs 8 and 9 shows the causes of traffic accidents. Different types of accidents due to collisions, violations of traffic rules and highways [28]. These are some of the old datasets which are taken from the government survey document to show how the traffic accidents took place and also shows the number of accidents, the person dies and the person injured from 2017 to 2018.

LHS	->	RHS	Confidence
Brain Injuries	->	Broken Bones	100.00%
Brain Injuries,Collision	->	Broken Bones	100.00%
Brain Injuries,Hit-Run	->	Broken Bones	100.00%
Broken Bones,Collision	->	Brain Injuries	100.00%
Broken Bones,Hit-Run	->	Brain Injuries	100.00%
Collision	->	Broken Bones	100.00%
Collision	->	Brain Injuries	100.00%
Collision	->	Brain Injuries,Broken Bones	100.00%
DrunknDrive	->	Broken Bones	100.00%
DrunknDrive,Skull & Maxillofacial	->	Broken Bones	100.00%
Hit-Run	->	Broken Bones	100.00%
Hit-Run	->	Brain Injuries	100.00%
Hit-Run	->	Brain Injuries,Broken Bones	100.00%
SportingAccident	->	Skull & Maxillofacial	100.00%

Execution Time: 124 milliseconds

Figure 7: Predicting severity patterns using the ECLAT algorithm.

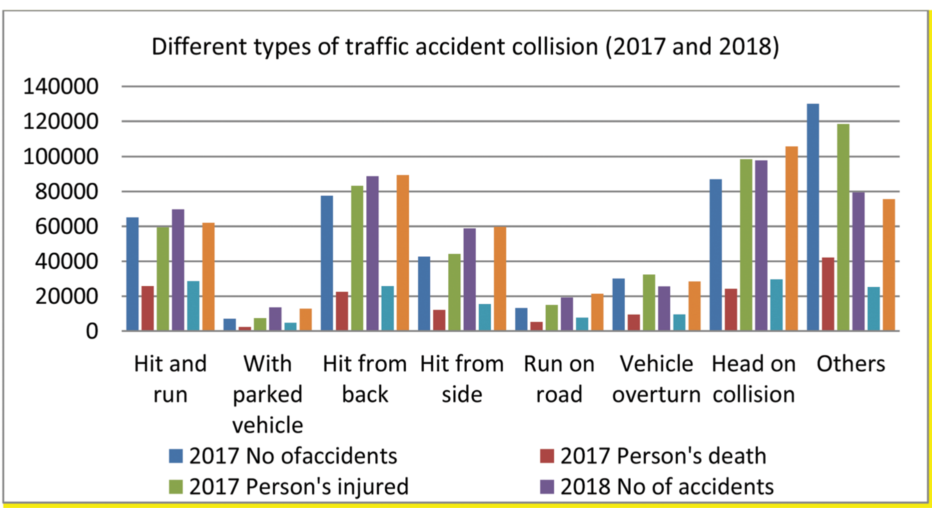


Figure 8: Different types of traffic accident collision of 2017 and 2018.

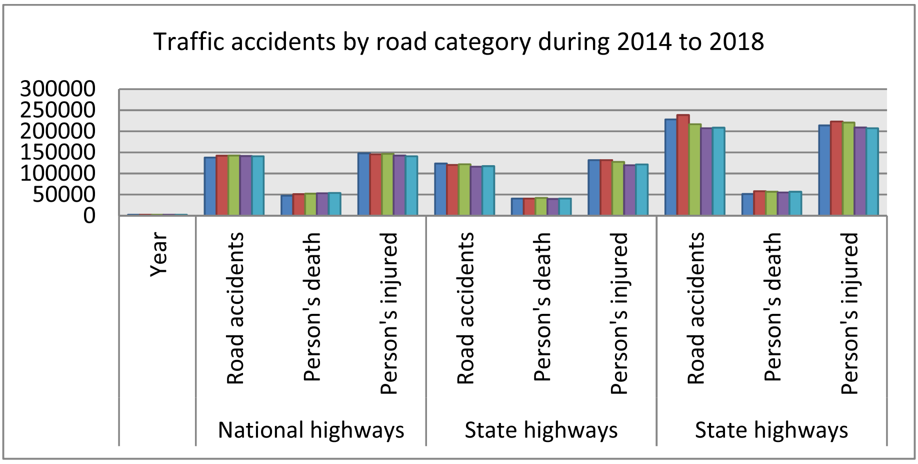


Figure 9: Traffic accidents due to traffic rules violations during 2017 and 2018.

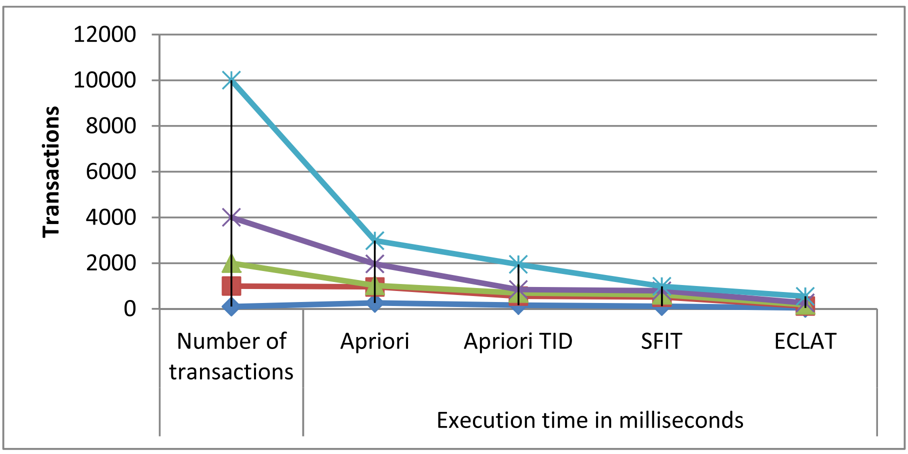


Figure 10: Time comparison between Apriori, Apriori TID, SFIT and ECLAT algorithm.

6 COMPARATIVE ANALYSIS OF PROPOSED METHOD

In this paper, we have done various experiments to scrutinize the performance of the proposed algorithms. The number of datasets 100, 1,000, 4,000 and 10,000 were taken to compare all the algorithms (Apriori, Apriori TID, SFIT and ECLAT) in terms of execution time represented in different colors.

Figure 10 shows the graphical representation of the execution time of all four algorithms. ECLAT algorithm is the best for finding out the frequent itemsets compared to other algorithms because the éclat algorithm took less time to discover the patterns of traffic accident severities from the given accident datasets.

7 CONCLUSION

Road safety represents a significant part of our lives, so it is necessary to reduce traffic accidents and their injury severities. The proposed work is a real-time application that is helpful for traffic departments as well as doctors to reduce traffic accidents and for providing proper treatments for injuries. In this paper, we used unsupervised learning methods to discover the patterns of traffic accidents and their injuries from the given datasets. The simulation results of the proposed methods for predicting traffic accident severities i.e., the Apriori algorithm predicts the patterns in 962 milliseconds, the Apriori TID algorithm predicts the pattern in 557 milliseconds, the SFIT algorithm predicts the pattern in 516 milliseconds and the ECLAT algorithm in predicts the pattern in 124 milliseconds. ÉCLAT algorithm is the best algorithm for discovering patterns of a traffic accident and its injury severities from the given traffic accident datasets, because it takes less time, cost and storage.

The proposed system in the future can be enhanced with a module like the Public notification module, Query Module and Predicting reasons for accidents. Public notification (SMS/ Email): We can add a public notification that helps people who met with an accident and they can get help to reach the hospital and get treatment. Query module: We can add a query module for the interaction between administrator and member (Traffic in charger or Doctor) to maintain more traffic accident records and the public can get better treatment for a particular predicted injury. We can also predict reasons for accidents which helps traffic departments to take precautionary measures.

ACKNOWLEDGEMENTS

No financial support was received to perform the research work of this manuscript.

REFERENCES

[1] Zhou, Z, Attention based stack resnet for citywide traffic accident prediction. *20th IEEE International Conference on Mobile Data Management (MDM)*, pp. 369-370, 2019.

[2] Amundsen, Finn Harald, and Guro Ranes, Studies on traffic accidents in Norwegian road tunnels. *Tunneling and Underground Space Technology,* pp. 3-11, 2000.

[3] Myounghoe Kim, Seongwon Lee, Jaehyun Lim, Jongeun Choi, and Seong Gu Kang, Unexpected Collision Avoidance Driving Strategy Using Deep Reinforcement Learning. *IEEE Access,* pp. 17243–17252, 2020.

[4] Wang, H., Huang, Y., Khajepour, A., Rasekhipour, Y., Zhang, Y. & Cao, D., Crash Mitigation in Motion Planning for Autonomous Vehicles. *IEEE Transactions on Intelligent Transportation Systems,* pp. 1–11. 2020.

[5] Albatayneh, O., Moomen, M., Farid, A., & Ksaibati, K., Complementary Modeling of Gravel Road Traffic-Generated Dust Levels Using Bayesian Regularization Feedforward Neural Networks and Binary Probit Regression. *International Journal of Pavement Research and Technology,* **6,** pp. 1-8, 2018.

[6] Fernandes, Fábio & Alves de Sousa, Ricardo & Ptak, Mariusz., Head Injury Simulation in Road Traffic Accidents. Springer, pp. 59-48, 2018.

[7] Li, G., Ma, H., Guan, T. *et al.,* Predicting Safer Vehicle Font-End Shapes for Pedestrian Lower Limb Protection via a Numerical Optimization Framework. *Int.J Automota Technol,* **21**, pp. 749–756, 2020.

[8] A., Fang, C., Qiu, L., Zhao and Y., Jin, Driver Risk Assessment Using Traffic Violation and Accident Data by Machine Learning Approaches. *3rd IEEE International Conference on Intelligent Transportation Engineering (ICITE)*, pp. 291-295, 2018.

[9] Y., He, Z., Liu, X., Zhou and B., Zhong, Analysis of Urban Traffic Accidents Features and Correlation with Traffic Congestion in Large-Scale Construction District. *International Conference on Smart Grid and Electrical Automation (ICSGEA)*, pp. 641-644, 2017.

[10] T., Jian, L., Zhi-qiang, X., Jian-feng and G., Hong-yu, Cause Analysis and Countermeasures of fatal Traffic Accidents on Road Passenger Transportation Based on Typical Cases. *25th International Conference on Transportation Information and Safety (ICTIS),* pp. 951-955, 2019.

[11] Masoumi, K., Forouzan, A., Barzegari, H., Asgari Darian, A., Rahim, F., Zohrevandi, B. & Nabi, S., Effective Factors in Severity of Traffic accident-related Traumas an Epidemiologic Study Based on the Haddon Matrix. *Emergency (Tehran, Iran)*, pp. 78–82, 2018.

[12] Tripathi, S., Shetty, U., Hasnain, A. & Hallikar, R., Cloud-Based Intelligent Traffic System to Implement Traffic Rules Violation Detection and Accident Detection Units. *3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 22-26, 2019.

[13] Wadhahi, N.T.S.A., Hussain, S.M., Yosof, K.M., Hussain, S.A., & Singh, A.V.,Accidents Detection and Prevention System to reduce Traffic Hazards using IR Sensors. *2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, pp. 254-257, 2018.

[14] Zhang, J., Li, Z., Pu, Z., & Xu, C., Comparing Prediction Performance for Crash Injury Severity among Various Machine Learning and Statistical Methods. *IEEE Access*, pp. 1–1, 2018.

[15] Guang Yu, Shauo lu, Speed and lane changing control optimization research on accident and congestion concentrated section of guang-shen freeway*, IEEE Access*, pp. 1-8, 2019.

[16] Galatioto, F., Catalano, M., Shaikh, N., McCormick, E., & Johnston, R., Advanced accident prediction models, and impacts assessment. *IET Intelligent Transport Systems*, **12**(9), pp. 1131–1141, 2018.

[17] M., Zheng, et al., Traffic Accident's Severity Prediction: A Deep-Learning Approach-Based CNN Network. *IEEE Access*, pp. 39897-39910, 2019.

[18] Leijdesdorff, H. A., Gillissen, S., Schipper, I. B., & Krijnen, P., Injury Pattern and Injury Severity of In-Hospital Deceased Road Traffic Accident Victims in The Netherlands: Dutch Road Traffic Accidents Fatalities. *World Journal of Surgery*, **6**, pp. 268-272, 2020.

[19] Zong, F., Chen, X., Tang, J., Yu, P. & Wu, T., Analyzing Traffic Crash Severity With Combination of Information Entropy and Bayesian Network. *IEEE Access,* pp. 63288–*63302,* 2019.

[20] Gao, Z., Pan, R., Yu, R. & Wang, X., Research on Automated Modeling Algorithm Using Association Rules for Traffic Accidents. *IEEE International Conference on Big Data and Smart Computing (BigComp),* pp. 1-6, 2019.

[21] Song, J., Wen, R. & Yan, W., Identification of Traffic Accident Clusters using Kulldorff's Space-Time Scan Statistics. *2018 IEEE International Conference on Big Data (Big Data),* pp. 1-6, 2018.

[22] Tai, W.-K., Wang, H.-C., Chiang, C.-Y., Chien, C.-Y., Lai, K. & Huang, T.-C., RTAIS: Road Traffic Accident Information System. *2018 IEEE 20th International Conference on High-Performance Computing and Communications*, pp. 1-8, 2017.

[23] Wenqi, L., Dongyu, L. & Menghua, Y., A model of traffic accident prediction based on convolutional neural network. *2017 2nd IEEE International Conference on Intelligent Transportation Engineering (ICITE),* pp. 67-72, 2017.

[24] Xia, X.-L., Nan, B. & Xu, C., Real-Time Traffic Accident Severity Prediction Using Data Mining Technologies. *2017 International Conference on Network and Information Systems for Computers (ICNISC),* pp. 1-8, 2017.

[25] Xu, Q., Zheng, J. & Sun, C., Digital modeling analysis of urban road traffic capacity under the condition of traffic accidents. *2017 29th Chinese Control And Decision Conference (CCDC),* pp. 10-12, 2017.

[26] Singh, Sanjay, Road Traffic Accidents in India: Issues and Challenges. *Transportation Research Procedia*. **25**, pp. 4712-4723, 2017.

[27] Sanjana, K.R., S., Lavanya, and Y., Bevish Jinila, An approach on automated rescue system with intelligent traffic lights for emergency services. *2015 International Conference on Innovations in Information, Embedded, and Communication Systems (ICIIECS)*, pp. 1-4, 2015.

[28] E., Abdullah and A., Emam, Traffic Accidents Analyzer Using Big Data. *2015 International Conference on Computational Science and Computational Intelligence (CSCI*), pp. 392-397, 2015.

[29] Alkandari, Abdulrahman Abdullah; Aljandal, Meshari, Theory of Dynamic Fuzzy Logic Traffic Light Integrated System with Accident Detection and Action. *IEEE 2015 Second International Conference on Computing Technology and Information Management (ICCTIM) - Johor*, pp. 62–68, 2015.

[30] Fu Huilin, Zhou Yucai, The Traffic Accident Prediction Based on Neural Network. *IEEE Access*, pp. 1-6, 2011.

[31] Hui, Chen; Xianghui, Wang; Xiqiang, Zhang; Shaoli, Zhang, The Evaluation of Chinese Urban Traffic Management System Application Based on Intelligent Traffic Control Technology. *7th International Conference on Intelligent Computation Technology and Automation*, pp. 791–795. 2014.

[32] Li, Jikuang Yang, Otte, D., Zhou Xuenong, Zhang Chao, Yu Xiaoping, Li Weiqiang, A Study on Motor-Scooter Accidents in China and Germany. *IEEE 2013 Fifth International Conference on Measuring Technology and Mechatronics Automation (ICMTMA 2013), pp.* 110–113, 2013.

[33] Sherif, Hossam M.; Shedid, M. Amer; Senbel, Samah A., Real-time traffic accident detection system using a wireless sensor network. *IEEE 2014 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR),* pp. 59–64, 2014.

[34] Xie, Fang, Feng, Qingdong, Research of effects of the accident on traffic flow characteristics. *IEEE 2013 International Conference on Mechatronic Sciences, Electric Engineering and Computer (MEC) - Shengyang, China Proceedings 2013 International Conference on Mechatronic Sciences,* pp. 1670–1674, 2013.