Improved Machine Learning Algorithmic Methods for Upgraded Traffic Solution

Veena RathnaAugesteelia
Research Scholar
School of computing sciences
VISTAS
Chennai ,India
veenajsph@gmail.com

Dr.K.Rohini
Associate Professor
School of Computing sciences VISTAS
Chennai, India
rrohini16@gmail.com

***Abstract*—Machine learning-based modern traffic enhancement systems are becoming increasingly popular due to their ability to analyze large amounts of traffic data and make real-time decisions. The number of vehicles on the road has dramatically increased every day in recent years, but sadly, the infrastructure of the roads and traffic systems has not kept up with this growth, leading to ineffective traffic management. This imbalance has not only increased the congestion and traffic bottlenecks on the roads but also led to stressful conditions which amplified road accidents. By utilizing cutting-edge technologies using machine learning Algorithms provided an enriched approach as the solution for traffic congestion and road accidents. In this evaluation of the ML-based traffic management approach various classification techniques such as decision tree, Random forest, Artificial neural network are implemented with Traffic dataset and the accuracy of these algorithms is compared with the proposed model. Among them the proposed obtained higher accuracy than the existing one. The solutions, together with the underlying technology, benefits, and limitations are need to be analyzed rigorously and frame an improved machine learning based traffic solution for avoiding congestion and accidents.**

In this research work a modified New Extra Trees Classifier algorithm using Python is proposed to predict the severity of an accident based on various attributes such as age band of driver, sex of driver, educational level, vehicle driver relation, driving experience, lanes or medians, types of junction, road surface type, light conditions, weather conditions, type of collision, vehicle movement, pedestrian movement, and cause of accident. This proposed work also focuses upon the ability to effectively forecast road traffic congestion as one of its key objectives. The road segment, Code, latitude, longitude, ward Name, alarm Type, speed, and recorded Time, etc.

These systems use various machine learning algorithms such as neural networks, decision trees, and random forests to predict traffic patterns, optimize traffic flow, and reduce congestion. Some of the specific applications of machine learning-based traffic management systems include the main focus on the Predicting the expected level of traffic using machine learning algorithms such as Extra Trees Classifier. The Real-time traffic management using image detection and machine learning algorithms to optimize traffic signal switching times. More intensive Deeper analysis of traffic data using machine learning algorithms to make traffic management systems smarter and more effective. Advanced traffic control systems using AI to optimize the flow of vehicles and reduce the number of accidents. Traffic management systems using machine learning algorithms to monitor vehicles, detect and prevent traffic congestion, and suggest alternative routes

These machine learning-based traffic enhancement systems have the potential to improve road safety, reduce traffic congestion, and optimize traffic flow. They can also be used to predict and prevent accidents, reduce travel time, and improve the overall efficiency of transportation systems.

Keywords— Machine Learning, Big data technology, Traffic systems, Traffic congestion, accidents, Extra Trees Classifier

I.

II. INTRODUCTION

As the world population continues to grow at an unprecedented rate, so does the demand for efficient and effective transportation systems. Traffic management has become an integral part of modern urban planning and development. Traditional traffic management techniques have been used for decades, but they have limitations, such as high costs and low accuracy rates. With the advent of new technologies, machine learning techniques have emerged as a promising solution to improve traffic management. In this article, will explore the emerging machine learning techniques in traffic management and their benefits.

Every three and a half minutes, by the Indian government's reckoning, one of its citizens is killed in a traffic accident. That adds up to 150,000 people a year. But in all likelihood, the carnage is much worse: The World Health Organization (WHO) thinks there are 300,000 traffic deaths in India a year—more than there are people in Barbados. That is not just a function of India's enormous population. China has roughly the same number of people, but its authorities count just 58,000 fatalities each year [1]. In fact, Indiahome to just 10% of the world's registered vehicles, accounts for 22% of traffic deaths, the World Bank estimates, using WHO data. It is a daily massacre, taking place in full public view but provoking little consternation. [2]

The maximum congestion (103 percent) in Bengaluru recorded on August 20, 2019. Delhi has the maximum number of cars. Bangalore has declared as the world's worst city in terms of traffic congestion.

Fig1. Photo of traffic road in Bangalore

According to the Global Traffic Index-2019, conducted by a Netherlands based firm Tom-Tom, people spent 243 hours in traffic jams in 2019. India has four such cities among the world's 10 highest traffic-congested cities.

According to the TomTom vehicle navigation company, about the annual traffic index that it has surveyed in 416 cities, in 57 countries and 6 continents. It is the ninth annual edition of TomTom which briefs the 10 highest traffic-congested cities as shown in the table1.

Top 10 cities with worst traffic in the world

City	**Congestion %**
Bengaluru, India	71%
Manila, Philippines	71%

Bogota, Colombia	68%
Mumbai, India	65%
Pune, India	59%
Moscow, Russia	59%
Lima, Peru	57%
Delhi, India	56%
Istanbul, Turkey	55%
Jakarta, Indonesia	53%

Table 1

Specific Traffic Findings in India

• According to the report, it takes 71% more time to complete the journey of 30 minutes. Not only this, four Indian cities have been included in the world's top 10 cities with most congested traffic.

• As per average, people in Bangalore spend 243 hours every year in traffic jams which is equal to 10 daysand3hours.

• The maximum congestion (103 percent) in Bengaluru was recorded on August 20, 2019, while the lowest congestion (30%) was recorded on April 06,2019.

• Similarly, people driving in Mumbai spend 209 hours in traffic every year.

• Maximum congestion was recorded on September 09, 2019 (101%) while the lowest congestion was recorded on March 21, 2019.

• According to the report, Delhi has the maximum number of cars while it is fourth on the list.

Global Findings

Manila (Philippines), Bogota (Colombia), Moscow (Russia), Lima (Peru), Istanbul (Turkey) and Jakarta (Indonesia) are also in the top-10 list. According to the report, people are spending an average of 193 hours of their life in traffic-jam every year. Greensboro-High Point of USA has the least traffic congestion in the world.[3]

Traffic Safety and Road Conditions in India

Travel by road in India is dangerous. India leads the world in traffic-related deaths and a number of U.S. citizens have suffered fatal traffic accidents in recent years. Extreme caution needs to be exercised when crossing streets, even in marked pedestrian areas, and only cars have seatbelts availability and usage. Sometimes Seat belts are not common in taxis.

Travel at night is particularly hazardous. Buses, patronized by hundreds of millions of Indians, are convenient in that they serve almost every city of any size. However, they are usually driven fast, recklessly, and without consideration for the rules of the road. Accidents are quite common. Trains are safer than buses, but train accidents still occur more frequently than in other countries. Our area of Buses and trucks often run red lights and merge directly into traffic at yield points and traffic circles.

The faster atpresent, thethree-dimensional traffic monitoring platform is completed or built in most of the cities. The pedestrian GPS, Vehicle GPS camera and other monitoring tools are not used to monitorroad cross-section flow, monitor vehicle travel speed and intersection shunt. The real-time evaluation of the operational status of urban roads is possible using these data. Huge amountof real-time monitoring data forms massive traffic information.It provides effective data for traffic jam prediction, and it also requires traffic jam forecast to proficiently and comprehensively cover the whole urban transportation network. In many places, traffic congestion has gotten worse, and traffic accidents have become more common. These issues created traffic management issues that needed to be addressed. The traditional method of traffic management appears to be inadequate in response to rise traffic request and the strain on transportation properties. To improve traffic management system the need of intelligent transportation system (ITS) has become increasingly important. To improve the efficiency of urban traffic network, the traditional road traffic management methods are deeply rehabilitated by using contemporary information technology, supported by machine learning and Big Data technology.

In Evaluation of this traffic stream and congestion the flow of traffic is a serious aspect that donates to traffic congestion. Traffic flow is referred as the amount of number of passing cars per unit time inone of the directions of a road is referred to as traffic flow. The number of vehicles travelling through a unit of time is during a specific direction at a traffic crossroads can be calculated and predicted using classification and regression with machine learning algorithms based on the traffic flow, speed and Density.

Traditional Traffic Management Techniques vs. Machine Learning Techniques

Traditional traffic management techniques, such as traffic lights and traffic signs, have been used for decades. They are reliable and effective in controlling traffic flow, but they have limitations. For example, traffic lights are programmed to switch at predefined intervals, regardless of the actual traffic conditions. This can lead to congestion and delays during peak hours.

Machine learning techniques, on the other hand, can adapt to changing traffic patterns and make real-time decisions based on the current traffic situation. For example, machine learning algorithms can adjust traffic signals based on the volume of traffic at a particular intersection. This can lead to a smoother flow of traffic and reduce congestion.

Machine Learning Algorithms Used in Traffic Management

There are several machine learning algorithms used in traffic management. One of the most commonly used algorithms is the neural network. Neural networks are designed to simulate the way the human brain works by processing and analyzing data in parallel. They are used to predict traffic patterns and make decisions based on the current traffic situation.

Another commonly used algorithm is the decision tree. Decision trees are used to make decisions based on a series of if-then statements. They are useful in traffic management for predicting traffic flow based on historical data.

Support vector machines (SVMs) are also used in traffic management. SVMs are used to classify data into different categories based on their attributes. They are used to predict traffic congestion and accidents.[5]

Benefits of Using Machine Learning in Traffic Management

There are several benefits of using machine learning in traffic management. First, machine learning algorithms can process vast amounts of data in real-time. This allows traffic managers to make informed decisions and take proactive measures to prevent congestion and accidents.

Second, machine learning algorithms can adapt to changing traffic patterns. They can adjust traffic signals and reroute traffic to avoid congestion. This leads to a smoother flow of traffic and reduces travel time.

Third, machine learning algorithms can improve safety on the roads. They can predict accidents and alert drivers to potential hazards. This can reduce the number of accidents and fatalities on the roads.[4]

Emerging Machine Learning Techniques in Traffic Management

Real-time Traffic Prediction: One of the emerging machine learning techniques in traffic management is real-time traffic prediction. Real-time traffic prediction uses machine learning algorithms to analyze current traffic data and predict future traffic conditions. This allows traffic managers to take proactive measures to prevent congestion and accidents.

Dynamic Traffic Control: Another emerging machine learning technique in traffic management is dynamic traffic control. Dynamic traffic control uses machine learning algorithms to adjust traffic signals based on the volume of traffic at a particular intersection. This leads to a smoother flow of traffic and reduces congestion.

Intelligent Traffic Signal Control: Intelligent traffic signal control is another emerging machine learning technique in traffic management. Intelligent traffic signal control uses machine learning algorithms to analyze traffic data and adjust traffic signals based on the current traffic situation. This leads to a more efficient flow of traffic and reduces travel time.

Difference between decision tree algorithm, random forest algorithm and extra trees classifier algorithm:

The decision tree algorithm, random forest algorithm, and extra trees classifier algorithm are all machine learning algorithms used for classification and regression tasks. But they differ in their approach to building models and making predictions.
A decision tree is a simple, yet powerful algorithm that builds a tree-like model of decisions and their possible consequences. It works by recursively splitting the data into subsets based on the values of the input features, and choosing the feature that maximizes the separation between the target classes. Once the tree is built, predictions are made by traversing the tree from the root to the leaf node that corresponds to the input data.[6]

Random forest is an ensemble learning algorithm that uses multiple decision trees to improve the accuracy and stability of predictions. It works by building a set of decision trees on random subsets of the data, and combining their outputs by taking a majority vote. This helps to reduce overfitting and increase the generalization performance of the model.

Extra trees classifier, also known as extremely randomized trees, is another ensemble learning algorithm that builds a forest of decision trees using random splits and random selection of features. It constructs multiple trees like RF algorithms during training time over the entire dataset.The main difference between random forest and extra trees is that extra trees use a more aggressive approach to split selection, where each node is split based on a random threshold instead of the optimal threshold. This makes extra trees faster to train than random forest, but may lead to a slightly lower accuracy.

Extra Trees and Random Forest handle overfitting

Both Extra Trees and Random Forest are ensemble learning algorithms that can handle overfitting. Overfitting occurs when the model is too complex and fits the training data too closely, resulting in poor generalization to new data. To avoid overfitting, both algorithms use techniques such as bootstrap sampling and aggregation to create multiple decision trees and reduce variance. Random Forest also uses feature bagging, which randomly selects a subset of features for each tree, to reduce correlation between trees and improve generalization.[7] Extra Trees, on the other hand, adds an additional level of randomization by choosing the split points randomly, which further reduces variance and makes the algorithm less influenced by certain features or patterns in the dataset. In summary, both Extra Trees and Random Forest use ensemble techniques to reduce overfitting and improve generalization.[8]

Advantages and Applications of Extra Trees Classifier:

The main advantage of Extra Trees is the reduction in bias. This is in terms of sampling from the entire dataset during the construction of the trees. Different subsets of the data may introduce different biases in the results obtained, hence Extra Trees prevents this by sampling the entire dataset.[9]

Another advantage of Extra Trees is that they reduce variance. This is a result of the randomized splitting of nodes within the decision trees, hence the algorithm is not heavily influenced by certain features or patterns in the dataset.Extra Trees are applied to perform three types of tasks such as classification and regression and feature selection.[10]

Proposed work:

In this research work proposes a NewExtra Trees Classifier, that has provided a best solution for traffic enhancement.The random selection of features leads to greater diversity among the decision trees and can result in better performance and reduced overfitting.

The Extra Trees Classifier also randomly selects cut-points for each feature, rather than using the median cut-point as in traditional decision trees. This further increases the diversity of the trees and can lead to improved performance.

The proposed NewExtra Trees Classifier can provide a more robust and accurate model for classification tasks. It is particularly useful when dealing with noisy or incomplete data, as well as when there are a large number of irrelevant features in the dataset.

This proposed algorithm is used to provide a best enhanced traffic warning system and accident severity prediction system.

Enhanced Traffic Warning System with NewExtra Trees Classifier:

The New Extra Trees Classifier algorithm to enhance traffic warning system with the following attributes: Code, latitude, longitude, wardName, alarmType, speed, and recordedTime:

1. Begin by randomly selecting a subset of features at each split point to create decision trees. The number of features to select can be set based on

prior knowledge or through cross-validation.

2. Train a set of decision trees using the training data, with each tree using a different subset of features and split points.

3. At each split point, randomly select a subset of cut-points for each feature to further increase the diversity of the trees.

4. Combine the predictions of all decision trees to predict the likelihood of an accident or traffic congestion. The classification can be based on a majority vote, or by using weighted voting based on the accuracy of each decision tree.

5. Evaluate the performance of the model using metrics such as accuracy, precision, recall, and F1 score.

6. If the model is not performing well, adjust the number of decision trees, the number of features, or the number of cut-points to improve performance.

7. Once the model is trained, it can be used to provide an enhanced traffic warning system based on the attributes such as Code, latitude, longitude, wardName, alarmType, speed, and recordedTime.

8. The algorithm can be used to predict the likelihood of an accident or traffic congestion based on the location, time, and other relevant attributes.

9. Based on the predictions, a warning system can be triggered to alert drivers or traffic management authorities to take appropriate measures to reduce accidents or alleviate traffic congestion.

10. The warning system can be integrated with other traffic management systems to provide real-time information to drivers and traffic management authorities.

By using this modified Extra Trees Classifier algorithm, traffic management authorities can enhance their traffic warning systems and reduce the likelihood of accidents and traffic congestion. The algorithm can handle noisy or incomplete data and a large number of irrelevant features, which makes it particularly useful in predicting the likelihood of an accident or traffic congestion based on various attributes

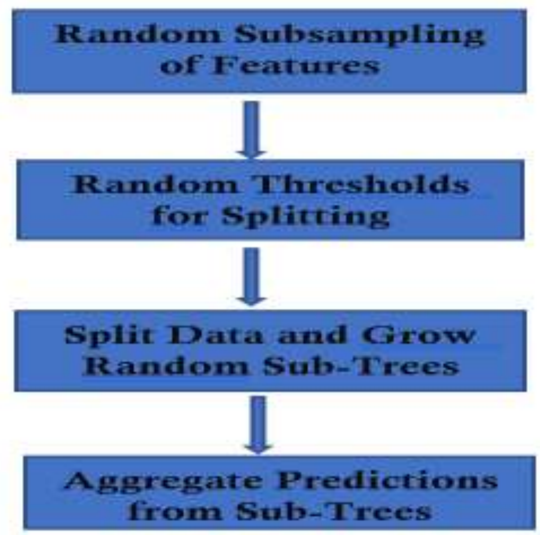


Fig1. steps in extra tree classifier

Description of dataset

Here two types of dataset are used are used for implementation and comparative Evaluation of the proposed New Extra tree classifier.

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 152276 entries, 0 to 207616
Data columns (total 7 columns):
deviceCode_deviceCode               152276 non-null int64
deviceCode_location_latitude        152276 non-null float64
deviceCode_location_longitude       152276 non-null float64
deviceCode_location_wardName        152276 non-null object
deviceCode_pyld_alarmType           152276 non-null object
deviceCode_pyld_speed               152276 non-null int64
deviceCode_time_recordedTime_$date  152276 non-null object
dtypes: float64(2), int64(2), object(3)
memory usage: 9.3+ MB
```

Fig 2. Number of Dataset entries and data Types of the Attributes

```
              Code   latitude  longitude        wardName  alarmType  \
0       8.645040e+14  12.984595  77.744087        Kadugodi        PCW
2       8.645040e+14  12.984595  77.744087        Kadugodi        PCW
4       8.645040e+14  12.987233  77.741119  Garudachar Playa        FCW
6       8.645040e+14  12.987233  77.741119  Garudachar Playa        FCW
8       8.645040e+14  12.987503  77.740051            Hudi  Overspeed
...              ...        ...        ...             ...        ...
415224  8.645040e+14  12.976435  77.741516        Kadugodi       UFCW
415226  8.645040e+14  12.986425  77.745117        Kadugodi       UFCW
415228  8.639770e+14  12.969396  77.749886         Hagadur  Overspeed
415230  8.639770e+14  12.974123  77.746841         Hagadur        FCW
415232  8.639770e+14  12.975480  77.744125         Hagadur        PCW

        speed             recordedTime
0        32.0  2018-02-01T01:48:59.000Z
2        32.0  2018-02-01T01:48:59.000Z
4        41.0  2018-02-01T01:50:00.000Z
6        41.0  2018-02-01T01:50:00.000Z
8        37.0  2018-02-01T01:50:11.000Z
...       ...                      ...
415224    0.0  2018-07-30T10:56:12.000Z
415226   12.0  2018-07-30T11:07:19.000Z
415228   17.0  2018-07-30T11:14:02.000Z
415230   36.0  2018-07-30T11:16:24.000Z
415232   28.0  2018-07-30T11:25:51.000Z

[207617 rows x 7 columns]
```

Fig:3 SAMPLE DATA

TRAIN DATA INFORMATION

- Preprocessing
  Data size 207617 After Preprocessing 152276
- Updated column names of train dataframe: Index(['deviceCode', 'latitude', 'longitude', 'wardName', 'alarmType', 'speed', 'recordedDateTime'],dtype='object')

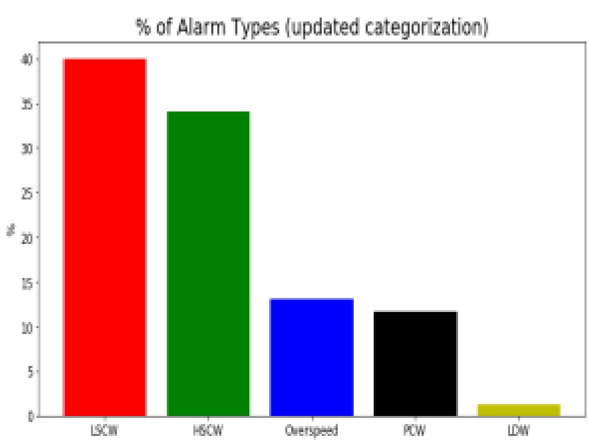


Fig.4The percentage of Different Alarms in the Dataset.

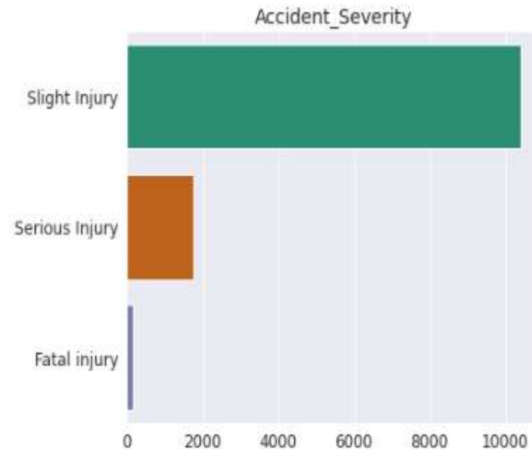


Fig5. Description of the Severity of Accident.

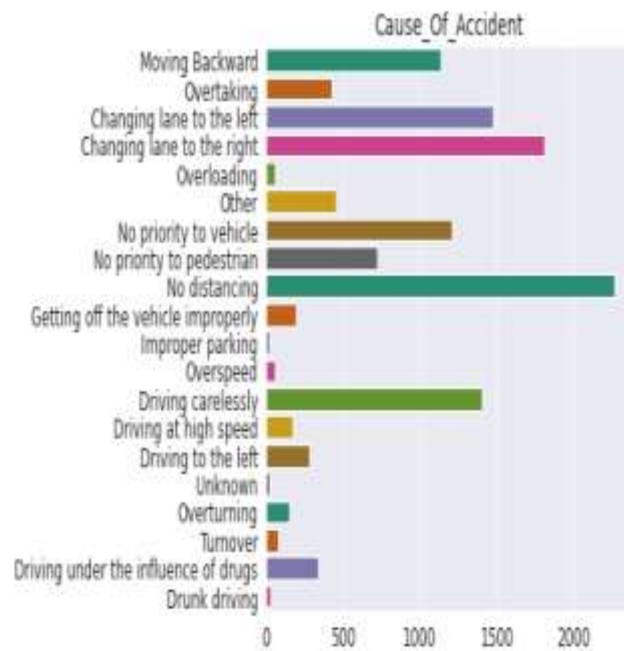


Fig.6 List of attributes Accident Dataset

EnrichedAccident Prediction System with NewExtra Trees Classifier:

A new modified New Extra tree classifier algorithm to predict the severity of an accident based on various attributes such as age band of driver, sex of driver, educational level, vehicle driver relation, driving experience, lanes or medians, types of junction, road surface type, light conditions, weather conditions, type of collision, vehicle movement, pedestrian movement, and cause of accident.

Steps of the proposed New Extra Trees Classifier algorithm to predict the severity and enhance the Road safety.

1. Begin by randomly selecting a subset of features at each split point to create decision trees. The number of features to select can be set based on prior knowledge or through cross-validation.

2. Train a set of decision trees using the training data, with each tree using a different subset of features and split points.

3. At each split point, randomly select a subset of cut-points for each feature to further increase the diversity of the trees.

4. Combine the predictions of all decision trees to classify the accident severity. The classification can be based on a majority vote, or by using weighted voting based on the accuracy of each decision tree.

5. Evaluate the performance of the model using metrics such as accuracy, precision, recall, and F1 score.

6. If the model is not performing well, adjust the number of decision trees, the number of features, or the number of cut-points to improve performance.

7. Once the model is trained, it can be used to predict the severity of an accident based on the attributes such as age band of driver, sex of driver, educational level, vehicle driver relation, driving experience, lanes or medians, types of junction, road surface type, light conditions, weather conditions, type of collision, vehicle movement, pedestrian movement, and cause of accident.

By using this New Extra Trees Classifier algorithm, traffic management authorities can predict the severity of accidents and take appropriate measures to reduce accidents and improve road safety. The algorithm can handle noisy or incomplete data and a large number of irrelevant features, which makes it particularly useful in predicting the severity of accidents based on various attributes.

```
confussion matrix
[[   1    5   46]
 [   0   65  487]
 [   0  175 2916]]


Train Accuracy of  ExtraTreesClassifier: 100.0

Test Accuracy of ExtraTreesClassifier: 80.70365358592693

              precision    recall  f1-score   support

           1       1.00      0.02      0.04        52
           2       0.27      0.12      0.16       552
           3       0.85      0.94      0.89      3091

    accuracy                           0.81      3695
   macro avg       0.70      0.36      0.36      3695
weighted avg       0.76      0.81      0.77      3695
```

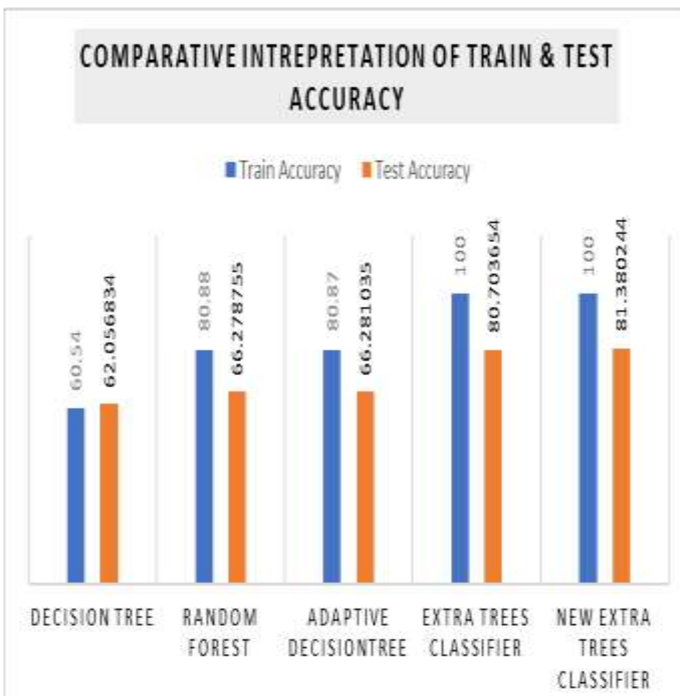
Fig.7 classification Accuracy with confusion matrixof Extra Trees Classifier

```
confussion matrix
[[   0    4   48]
 [   0   63  489]
 [   0  147 2944]]


Train Accuracy of  NewExtraTreesClassifier: 100.0

Test Accuracy of NewExtraTreesClassifier: 81.38024357239513

              precision    recall  f1-score   support

           1       0.00      0.00      0.00        52
           2       0.29      0.11      0.16       552
           3       0.85      0.95      0.90      3091

    accuracy                           0.81      3695
   macro avg       0.38      0.36      0.35      3695
weighted avg       0.75      0.81      0.77      3695
```

Fig.8 classification Accuracy with confusion matrix of New Extra Trees Classifier

Model	Train Accuracy	Test Accuracy
Decision Tree	60.54	62.056834
Random Forest	80.88	66.278755
Adaptive DecisionTree	80.87	66.281035
Extra Trees Classifier	100	80.703654
New Extra Trees Classifier	100	81.380244

Table 2. Contains the Train and Test Accuracy of Different existing and Proposed Models.



Graph.1 **Comparative Intrepretationof Train & Test Accuracy**

V.CONCLUSION

Machine learning techniques have emerged as a promising solution to improve traffic management. They can process vast amounts of data in real-time and make informed decisions based on the current traffic situation. By analyzing traffic data from various sources, machine learning algorithms able to learn traffic patterns and make predictions about future traffic conditions. The proposed New extra treeclassifier provides enriched solutions for both Traffic Congestion and road Accidents. The proposed work has been implemented and evaluated with other existing and previously proposed Adaptive decision tree algorithm. Although the Adaptive decision tree algorithm has produced the classification accuracy almost same as the Random Forest Algorithm. But the proposed New extra tree classifier proves to have the better improved the classification accuracy with 81.38. This proposed work allowsto enrich and enhance traffic management solution and paves way to take proactive measures to prevent congestion and accidents. The future of traffic enhancement looks promising with the emergence of machine learning techniques.

VI.REFERENCES

1. “Rear guard action: on car accident-related deaths". The Hindu. 6 September 2022. Retrieved 6 September 2022.
2. Data from World Health Organization Estimated Deaths 2012.
3. "Two-wheelers claimed highest number of lives in accidents in 2021: NCRB report". The Hindu. PTI. 30 August 2022. Retrieved 6 September 2022.
4. ZhenhuaWang, YangsenYu,DangchenJu,“Analysis and Prediction of Urban Traffic Congestion Based on Big Data”, International Journal on Data Science and Technology, October 2018.
5. Hua-pu Lu, Zhi-yuan Sun and Wen-cong Qu, “Big Data-Driven Based Real-Time Traffic Flow State Identification and Prediction”, Discrete Dynamics in Nature and Society, Volume2015.
6. Sachin Kumar, Durga Toshniwal, “A data mining framework to analyze road accident data”, Journal of Big Data, a Springer Open Journal,2015.
7. Sung, Hong Ki and Chong, Kyu Soo, “Development of Road Traffic Analysis Platform Using Big Data”, International Conference on Advances in Big Data Analytics,ABDA’17.
8. Krithika, D. R., and K. Rohini. "Ensemble Based Prediction of Cardiovascular Disease Using Bigdata analytics." *2021 International Conference on Computing Sciences (ICCS)*. IEEE, 2021.
9. Ramya, A., and K. Rohini. "Comparative evaluation

of machine learning classifiers with Obesity dataset." *2021 International Conference on Computing Sciences (ICCS)*. IEEE Computer Society, 2021

10. Liu, Yunxian,Wu, Hao2017/12/01 “Prediction of Road Traffic Congestion Based on Random Forest” in the conference 2017 10th International Symposium on Computational Intelligence and Design(ISCID), DOI 10.1109/ISCID.2017.216.