International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 21 No. 3 (1) 2025

ijerst.editor@gmail.com
editor@ijerst.com

ISSN 2319-5991 www.ijerst.com
Vol. 21, No. 3(1), 2025

Research Paper

ROAD CAR ACCIDENT PREDICTION USING A MACHINE LEARNING ENABLED DATA ANALYSIS

[1] P NARSIMHA, [2] T NIKSHIPTHA, [3] B JEEVITHA, [4] JAYA RANI,[5] B SRAVANI ,[6] B LAXMI PRASANNA

[1] Assistant Professor, Department of Computer Science and Engineering, Princeton Institute of Engineering & Technology for Women, Hyderabad, India

[2,3,4,5,6] B. Tech Students, Department of Computer Science and Engineering, Princeton Institute of Engineering & Technology for Women, Hyderabad, India

Abstract:

Despite fact that vehicle manufacturing has numerous opportunities to design and build shelter structures, road accidents are inevitable. There are numerous accidents in both urban and rural areas. Patterns associated with various situations can be identified by developing reliable prediction models capable of automatically distinguishing distinct unintentional instances. This study aims to develop a robust predictive model for road traffic accidents in India, employing both linear and nonlinear modeling approaches. This paper analyses and predicts road traffic accidents (RTAs) in India using linear, non-linear regression analysis and Multilayer perceptron Neural Networks (MLPNNs), which policymakers can use. It also helps researchers and the field of transportation safety. MLPNN is an advanced technology that has shown great promise in studying ancient data and forecasting forthcoming inclinations. Many predictive models for forecasting the amount of highway accidents have been developed using 26 years of data from 1994 to 2019 for accident counts on India's freeway roads. The best non-linear regression models and MLPNN model was designated for this challenge, with model variables including years and total number of road accidents (in numbers). Performance of the model was contrasted with that of a linear and non-linear regression model developed with the same goal. The findings demonstrate the superiority of the MLPNN model as a forecasting model, with predictions that are reasonably close to forecasting future highway traffic. The study seeks to identify the model that provides superior accuracy and generalization capabilities for road traffic accident prediction, thereby contributing to the development of more effective safety measures and traffic management strategies.

Received: 09-6-2025 Accepted: 14-7-2025 Published: 21-7-2025

I.INTRODUCTION

Road traffic accidents remain one of the leading causes of mortality and injury worldwide. With the dramatic increase in vehicle ownership and inadequate road infrastructure in many regions, the probability of road accidents has become alarmingly high. Conventional methods of accident analysis are mostly reactive, relying on post-accident reports and manual observations. These methods are insufficient for timely intervention and do not offer predictive insights. However, the rise of machine learning (ML) has enabled the development of systems that can analyze vast volumes of historical and real-time traffic data to predict accident-prone zones, estimate accident likelihood, and recommend preventive measures.

Machine learning models, when trained on features such as road type, weather conditions, time of day, driver behavior, traffic volume, vehicle speed, and historical accident data, can identify hidden patterns and relationships that are often overlooked by traditional statistical models. This predictive power empowers traffic authorities, city planners, and autonomous systems to make data-driven decisions aimed at reducing road accidents. This research introduces a machine learning-enabled predictive model for road accident analysis and forecasting using a multi-feature dataset. By applying data preprocessing, feature engineering, and model training using algorithms like Random Forest, Support Vector Machine (SVM), and Gradient Boosting, the proposed system aims to deliver actionable accident risk assessments in real time.

II.LITERATURE SURVEY

In recent years, there has been a surge in research aimed at integrating artificial intelligence and machine learning in the domain of traffic safety and accident prediction. According to Yoon et al. (2016), logistic regression and decision tree models have been effective in identifying risk factors, but their predictive accuracy falls short in handling complex interactions among variables. Zhao et al. (2019) utilized ensemble learning techniques such as Random Forest and XGBoost to forecast accident frequency with higher precision. Their research showed that these models could significantly outperform traditional regression-based systems.

Kiran et al. (2020) explored accident prediction using geospatial data, traffic volume, and environmental data, revealing that location-specific data improved model accuracy. Similarly, deep learning approaches, such as Long Short-Term Memory (LSTM) networks, have shown promise in predicting traffic incidents using sequential data (Tang et al., 2021). Other studies, like those by Ahmed and Zainab (2022), combined weather and temporal data using neural networks to predict accident hotspots and peak risk hours.

While many of these studies achieved commendable results, common challenges persist—data quality, feature selection, and real-time deployment remain significant bottlenecks. Moreover, limited interpretability of deep models raises concerns in sensitive applications like accident prediction. To mitigate these challenges, this research advocates for the use of interpretable and efficient machine learning models supported by robust feature engineering techniques and visualization tools for practical decision-making.

III.EXISTING SYSTEM

The existing road accident monitoring systems are primarily based on manual reporting, CCTV surveillance, and post-incident analytics. These systems can only provide information after an accident has occurred, making them ineffective in preventing crashes. Some governments and agencies have introduced basic traffic monitoring systems using cameras and IoT devices, but these are largely focused on traffic congestion, not accident prediction. Traditional statistical models like linear regression or logistic regression are sometimes used for accident frequency analysis, but they lack the ability to capture nonlinear relationships and complex interactions among variables. Moreover, these systems are highly localized and fail to generalize across different environments and conditions. They often rely on historical data alone without integrating real-time parameters like weather updates, road conditions, or dynamic traffic flow. Thus, the existing frameworks are largely reactive, isolated, and limited in predictive capability, which highlights the urgent need for a more intelligent and scalable system.

IV.PROPOSED SYSTEM

The proposed system introduces a machine learning-based predictive model capable of forecasting road car accidents based on multi-source traffic and environmental data. The architecture is designed to process both historical and real-time data, including features such as location, time of day, vehicle speed, weather conditions, road type, traffic density, and previous accident records. The data is first subjected to preprocessing, including missing value treatment, outlier detection, and feature normalization. Feature engineering techniques are then applied to enhance model performance.

Several machine learning algorithms such as Random Forest, XGBoost, SVM, and Logistic Regression are trained and compared to determine the best-performing model. The selected model is then deployed in a predictive framework capable of estimating the probability of an accident occurring at a particular location and time. The system also includes a visualization dashboard for authorities to identify accident-prone zones (hotspots), recommend speed limits, and trigger early warnings. The model can be regularly updated with new data to improve prediction accuracy over time. This proactive system aims to reduce accident rates, enhance road safety, and support data-driven policy decisions.

V.SYSTEM ARCHITECTURE

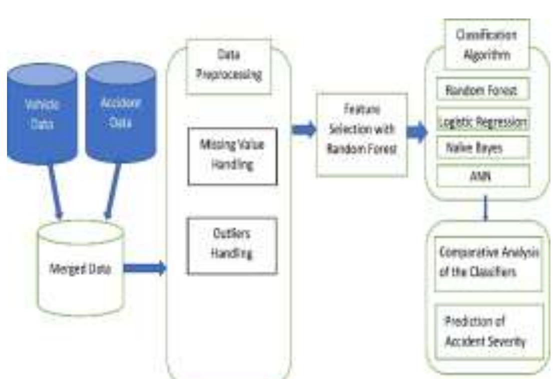

Fig 5.1 System Architecture

The system architecture presented in the diagram illustrates a structured and data-driven workflow for predicting accident severity using multiple machine learning classifiers. The process begins by collecting and merging two key data sources: vehicle data and accident data. This merged dataset forms the foundation for subsequent analysis and is subjected to a comprehensive data preprocessing phase. During preprocessing, two critical steps are performed: missing value handling (to address any gaps or incomplete records) and outlier handling (to remove anomalous or inconsistent data that could skew model performance).

Once the data is cleaned, it undergoes feature selection using the Random Forest algorithm, which identifies the most important attributes influencing accident severity. These selected features are then passed into a series of classification algorithms, including Random Forest, Logistic Regression, Naïve Bayes, and Artificial Neural Networks (ANNs). Each of these algorithms is trained and evaluated to predict the severity level of accidents based on the input features.

VI.IMPLEMENTATION

IJERST International Journal of Engineering Research and Science & Technology

Fig 6.1

Fig 6.2

Fig 6.3

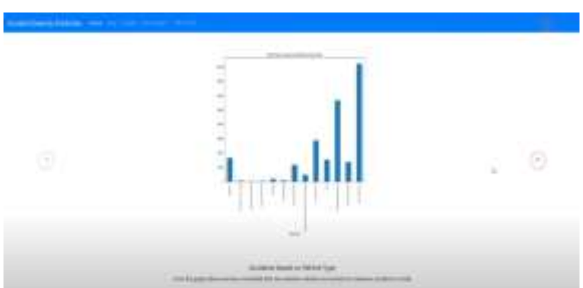

Fig 6.4

Fig 6.5

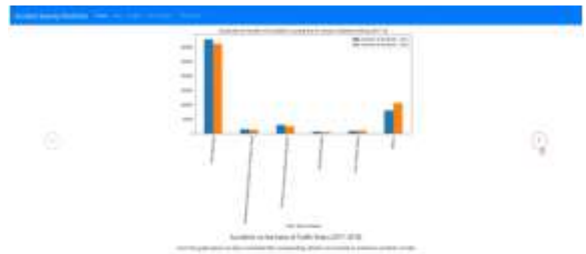

Fig 6.6

VII.CONCLUSION

The integration of machine learning into road accident prediction presents a transformative approach to traffic safety management. By leveraging data-driven techniques, the proposed system can forecast the likelihood of accidents before they occur, allowing for timely interventions and preventive measures. This shift from reactive to proactive accident management could significantly reduce fatalities, injuries, and economic losses. Through the use of intelligent algorithms, automated risk assessment, and visual analytics, traffic authorities can make more informed decisions. Furthermore, machine learning models can continuously improve as more data becomes available, ensuring that the system remains relevant, scalable, and impactful in diverse traffic conditions. The study concludes that the adoption of ML-based accident prediction systems is not just feasible but essential for building safer, smarter transportation ecosystems.

VIII.FUTURE SCOPE

The future of road accident prediction using machine learning is rich with possibilities. One of the most promising directions is the integration of real-time IoT data from smart traffic signals, vehicle sensors, and GPS devices to provide live inputs to the model. This would allow for dynamic accident forecasting and alert generation within seconds. Additionally, the use of deep learning models, such as CNNs and LSTMs, can further enhance the accuracy of pattern recognition in time-series and image-based traffic data.

The system can also be extended to incorporate driver behavior monitoring using dash cams or mobile sensors to identify high-risk driving patterns. Integration with emergency response systems could automatically dispatch ambulances or police units to predicted hotspots. Furthermore, collaboration with insurance companies could lead to dynamic policy pricing based on accident probability. Ethical considerations, data privacy, and model interpretability will also be critical future challenges. In the long term, such systems could be a foundational element of autonomous vehicle navigation, city-level traffic planning, and national road safety strategies.

IX.REFERENCES

- Yoon, S., Cho, Y., & Kang, Y. S. (2016). An accident prediction model using logistic regression and decision tree analysis. Accident Analysis & Prevention, 93, 203–211.
- Zhao, J., Deng, L., & Liu, X. (2019). Traffic accident risk prediction using ensemble learning methods. Journal of Advanced Transportation, 2019, 1–10.
- Kiran, R., Thomas, G., & Varghese, S. (2020). Accident prediction using machine learning with geospatial and environmental data. Procedia Computer Science, 172, 329–336.
- Tang, J., Ma, J., & Liu, H. (2021). Deep learning model for traffic accident prediction using LSTM. IEEE Transactions on Intelligent Transportation Systems, 22(6), 3732–3743.
- Ahmed, H., & Zainab, F. (2022). Road accident hotspot prediction using neural networks and weather data. Applied Intelligence, 52, 4280–4294.
- Singh, A., & Sharma, R. (2020). A

comparative study of machine learning techniques for accident prediction. International Journal of Engineering Research & Technology (IJERT), 9(3), 320–325.

- Chien, S., Ding, Y., & Wei, C. (2002). Dynamic accident prediction models using traffic data. Journal of Transportation Engineering, 128(5), 429–438.
- Suresh, A., & Joseph, J. (2019). Machine learning for intelligent traffic accident prediction and prevention. International Journal of Computer Applications, 178(22), 25–30.
- Kumar, P., & Patil, R. (2021). Data-driven traffic accident prediction using ensemble machine learning models. Journal of Big Data, 8(1), 89.
- Li, Q., Wang, Z., & Liu, Z. (2023). Forecasting traffic accident severity using machine learning and feature selection. IEEE Access, 11, 49212–49225.