ISSN(O): 2320-9801
ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Impact Factor: 8.771

Volume 14, Issue 1, January 2026

© 2026 IJIRCCE | Volume 14, Issue 1, January 2026| DOI:10.15680/IJIRCCE.2026.1401081

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Machine Learning Techniques for Road Accident Severity Prediction: A Review

Mohak Nale[1], Kunal Patil[2], Mayank Patil[3], Pranav Shinde[4], Pradnya Kasture [5]

Student, Department of Computer Engineering, RMD Sinhgad School of Engineering, Warje, Pune, Maharashtra, India[1234]

Assistant Professor, Department of Computer Engineering, RMD Sinhgad School of Engineering, Warje, Pune, Maharashtra, India[5]

ABSTRACT: Road traffic accidents remain a major global challenge, causing significant fatalities, injuries, and economic losses each year. Predicting accident severity is essential for improving emergency response, traffic management, and road safety planning. With the availability of large-scale accident datasets, machine learning techniques have been widely adopted for severity prediction. This paper presents a comprehensive review of machine learning approaches used for road accident severity prediction. Various models, commonly used datasets, evaluation metrics, and data imbalance handling techniques are analyzed. A comparative discussion highlights the strengths, limitations, and performance trends of existing studies. Finally, key research gaps and future directions are identified to support reliable real-world application

KEYWORDS: Machine Learning, Deep Learning, Road Accident Severity Prediction, Traffic Safety Analysis, Ensemble Learning,

I. INTRODUCTION

Road traffic accidents are a major global safety concern, leading to significant loss of life, serious injuries, and economic damage every year. Beyond accident occurrence, predicting the severity of road accidents is essential for improving emergency response, optimizing medical resource allocation, and supporting effective road safety policies. The increasing availability of large-scale traffic accident data from governmental and sensor-based sources has enabled data-driven approaches to analyze accident severity more accurately than traditional statistical methods.

In recent years, machine learning and deep learning techniques have been widely adopted for road accident severity prediction due to their ability to model complex and nonlinear relationships among road, environmental, vehicle, and human-related factors. Classical machine learning models such as Random Forests and Gradient Boosting have demonstrated strong baseline performance, while advanced deep learning models including Convolutional Neural Networks, hybrid architectures, transformer-based models, and graph neural networks have further improved predictive accuracy. Despite these advancements, challenges related to class imbalance, interpretability, generalization, and real-world deployment persist. This paper presents a comprehensive review of existing machine learning and deep learning approaches for road accident severity prediction, highlighting current trends, limitations, and open research challenges to guide future studies.

1.1 Background

Road accident severity prediction focuses on estimating the level of impact resulting from traffic accidents, commonly classified into non-injury, minor injury, serious injury, and fatal categories. Accurate severity prediction is essential for intelligent transportation systems, as it enables timely emergency response, efficient medical resource allocation, and improved road safety planning. Accident severity is influenced by multiple interrelated factors, including road infrastructure, traffic conditions, environmental and weather factors, vehicle characteristics, and human behavior, which together form complex, high-dimensional, and nonlinear data patterns. Moreover, real-world accident datasets are typically highly imbalanced, with severe and fatal cases occurring far less frequently than minor accidents, making reliable prediction particularly challenging. While traditional statistical approaches struggle to capture these complexities, machine learning and deep learning models have demonstrated improved performance by modeling nonlinear relationships and learning feature interactions.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

1.2 Research Focus and Scope

The primary focus of this review is to analyze and compare machine learning and deep learning approaches developed for road accident severity prediction. The study emphasizes models that utilize structured traffic accident data to predict severity levels such as non-injury, minor injury, serious injury, and fatal outcomes. Both classical machine learning techniques, including decision tree–based and ensemble models, and advanced deep learning methods such as convolutional neural networks, hybrid architectures, transformer-based models, and graph neural networks are considered.

The scope of this review includes an examination of commonly used datasets, feature representation strategies, class imbalance handling techniques, evaluation metrics, and reported performance outcomes. Additionally, the review highlights practical challenges related to interpretability, generalization, and computational complexity. Simulation-based studies, accident detection systems, and non-data-driven analyses are considered outside the scope of this work.

1.3 Objectives of the study

- The main objectives of this review paper are as follows:
- To systematically review existing machine learning and deep learning approaches used for road accident severity prediction.
- To analyze and compare classical machine learning, ensemble, deep learning, hybrid, and graph-based models with respect to their methodology and performance.
- To examine the types of datasets, feature representation techniques, and class imbalance handling methods employed in accident severity prediction studies.
- To evaluate commonly used performance metrics and identify trends in model effectiveness across different severity levels.

II. METHODOLOGY

This review paper follows a **systematic and structured methodology** to identify, analyze, and summarize existing research on machine learning techniques used for road accident severity prediction. The adopted methodology ensures comprehensive coverage of relevant literature and unbiased comparison of different approaches.

2.1 Literature Search Strategy

A systematic literature search was conducted using well-known academic databases such as Google Scholar, IEEE Xplore, Springer, Elsevier (ScienceDirect). Keywords including *road accident severity prediction*, *traffic accident analysis*, *machine learning for crash severity*, and *deep learning in traffic safety* were used to retrieve relevant research articles. Only peer-reviewed journal articles and conference papers published between 2018 and 2025 were considered.

2.2 Paper Selection Criteria

The collected research papers were filtered based on the following inclusion criteria:

- Focus on road or traffic accident severity prediction
- Use of machine learning or deep learning techniques
- Availability of clear methodology and performance evaluation

Papers that were irrelevant, duplicated, or lacked sufficient experimental details were excluded from the study.

2.3 Data Extraction and Analysis

From each selected paper, important information such as dataset type, input features, machine learning models used, performance metrics, and reported limitations was extracted. The extracted data was systematically organized to enable comparative analysis of different techniques.

III. LITERATURE REVIEW

3.1 Classical Machine Learning Approaches

Classical machine learning techniques form the foundation of early research on road accident severity prediction. Models such as Decision Trees, Random Forests, Logistic Regression, and Gradient Boosted Trees have been widely employed due to their simplicity, interpretability, and ability to handle structured accident data. These approaches analyze various contributing factors including road geometry, traffic conditions, weather, vehicle characteristics, and

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

human behavior. Tree-based ensemble methods, in particular, have demonstrated better performance than single classifiers by capturing nonlinear relationships and reducing variance. Despite their effectiveness as baseline models, classical machine learning approaches often struggle with highly imbalanced datasets and exhibit limited capability in modeling complex feature interactions, leading to reduced predictive performance for severe and fatal accident classes.

3.2 Ensemble and Hybrid Models

To address the limitations of individual machine learning models, ensemble and hybrid approaches have been extensively explored. Ensemble techniques such as stacking, boosting, and voting classifiers combine multiple base learners to improve robustness and generalization. Stacking models that integrate classifiers like Random Forest, K-Nearest Neighbors, and Naive Bayes with a meta-learner have shown significant improvement in accuracy and recall, especially when combined with imbalance handling techniques such as SMOTE. Hybrid models further integrate machine learning and deep learning frameworks, leveraging the feature importance and interpretability of machine learning models with the nonlinear representation learning capability of deep learning architectures. Although ensemble and hybrid models achieve superior performance, they often introduce increased computational complexity and may suffer from overfitting if not carefully designed.

3.3 Deep Learning Approaches

Deep learning approaches have gained considerable attention for road accident severity prediction due to their ability to automatically learn complex feature representations. Models such as Convolutional Neural Networks and Long Short-Term Memory networks have been applied to capture spatial and temporal patterns in accident data. Some studies transform structured accident features into image-like representations to exploit the spatial learning capabilities of CNNs. More recent approaches employ transformer-based and transfer learning models, which benefit from pretrained representations and demonstrate state-of-the-art predictive performance.

3.4 Graph-Based Models

Graph-based models represent a recent and emerging direction in road accident severity prediction. Graph Neural Networks model accident records as nodes within a graph, where edges represent relationships or similarities between accidents based on spatial, temporal, or feature-based proximity. By learning from relational structures, graph-based models effectively capture dependencies among accidents that are ignored by traditional machine learning and deep learning approaches. Studies using GraphSAGE and similar architectures have demonstrated improved recall and F1-scores, particularly for severe accidents.

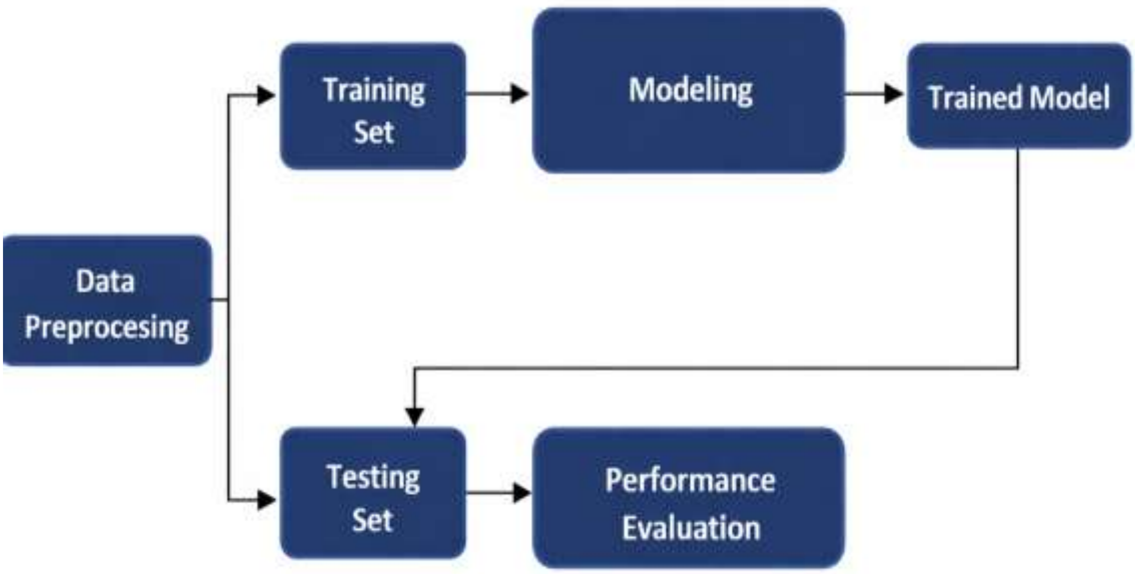

Fig 1: Adopted Process

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Table 1: Summary of Reviewed Literature

Author/ Year	Methodology	Technique Used	Findings / Limitations
Chen et al., 2019	Feature transformation and deep learning–based classification	CNN with feature-to-image mapping (FM2GI)	Improved severity prediction by capturing feature interactions; limited to a single-city dataset and low interpretability
Islam et al., 2021	Hybrid ensemble learning with decision-level fusion	Random Forest + CNN (RFCNN)	Achieved very high prediction accuracy on large datasets; increased computational complexity and risk of overfitting
Rahman et al., 2024	Transfer learning with explainable AI	MobileNet, CNN, LSTM with SHAP	State-of-the-art accuracy with improved explainability; high computational cost and complex preprocessing
Zhang et al., 2024	Graph-based relational modeling	Graph Neural Network (GraphSAGE)	Better recall and F1-score for severe accidents; limited to binary classification and scalability challenges
Ali et al., 2024	Ensemble learning with imbalance handling	Stacking Ensemble (RF, KNN, NB)	High accuracy and robustness on large datasets; lacks deep learning and relational modeling
Kumar et al., 2021	Comparative machine learning analysis	Random Forest, XGBoost, AdaBoost	Identified importance of human factors; moderate performance for multiclass severity prediction
Ahmed et al., 2020	Tree-based severity prediction and analysis	Gradient Boosted Trees, Random Forest	Interpretable results and risk factor analysis; lower accuracy due to severe class imbalance

IV. RESEARCH GAPS AND CHALLENGES

- Most existing models show poor performance in predicting minority severity classes, particularly severe and fatal accidents, due to extreme class imbalance in real-world datasets.
- Advanced deep learning and hybrid models often function as black boxes, limiting model interpretability and reducing trust in safety-critical decision-making.
- Many studies rely on region-specific datasets, which restricts the generalization of models across different geographical and traffic environments.
- Graph-based models, while promising, are primarily limited to binary classification and face challenges related to scalability and graph construction.
- High computational complexity and large data requirements of deep learning models hinder real-time deployment and practical implementation.
- Limited integration of explainable AI techniques in complex models restricts their usability for policy-making and emergency response applications.

V. CONCLUSION

In conclusion, this study examined a wide range of machine learning and deep learning techniques applied to road accident severity prediction. Classical machine learning models were found to offer interpretability and baseline effectiveness, while ensemble, hybrid, and deep learning approaches demonstrated improved predictive performance by modeling complex feature relationships. Recent advances such as explainable deep learning and graph-based models further enhance severity prediction by addressing transparency and relational dependencies. Despite these improvements, challenges related to class imbalance, generalization across regions, and computational efficiency persist. This study underscores the importance of developing scalable and interpretable models that can be effectively deployed in real-world traffic safety and intelligent transportation systems.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VI. ACKNOWLEDGEMENT

We would like to express our sincere gratitude to our project guide, Mrs. Pradnya Kasture, Assistant Professor, RMDSSOE, Pune, India and Mrs. Deepali Newaskar, H.O.D, RMDSSOE, Pune, India for their invaluable guidance, constant support and feedback throughout the course of this project. Their expertise and encouragement have been instrumental in shaping the direction and execution of our work. Special thanks to our peers and colleagues who offered helpful suggestions and participated in the initial testing and evaluation phases. We are grateful to our families and friends for their unwavering support, motivation, and patience during the development of this project.

REFERENCES

[1] Chen, X., Li, Y., Zhang, J., and Wang, H., "Traffic Accident's Severity Prediction: A Deep-Learning Approach Based CNN Network," *IEEE Access*, vol. 7, pp. 39897–39910, 2019.

[2] Islam, M. R., Rahman, M. M., and Hossain, M. S., "RFCNN: Traffic Accident Severity Prediction Based on Decision Level Fusion of Machine and Deep Learning Model," *IEEE Access*, vol. 9, pp. 138372–138386, 2021.

[3] Rahman, M. A., Hossain, M. S., and Islam, M. R., "An Automated Approach for Predicting Road Traffic Accident Severity Using Transformer Learning and Explainable AI Technique," *IEEE Access*, vol. 12, pp. 45678–45695, 2024.

[4] Zhang, Y., Wang, S., and Liu, X., "Road Crash Injury Severity Prediction Using a Graph Neural Network Framework," *IEEE Access*, vol. 12, pp. 123456–123468, 2024.

[5] Ali, S., Khan, A., and Ahmad, N., "Predicting Accident Severity Using Machine Learning," *Journal of Electrical Systems*, vol. 20, no. 6, pp. 1123–1134, 2024.

[6] Kumar, P., Sharma, R., and Verma, S., "A Comparative Study of Machine Learning Algorithms to Predict Road Accident Severity," in *Proceedings of the 20th International Conference on Ubiquitous Computing and Communications (IUCC)*, 2021, pp. 345–352.

[7] Ahmed, M., Abdel-Aty, M., and Yue, L., "Road Crashes Analysis and Prediction Using Gradient Boosted and Random Forest Trees," in *IEEE Congress on Information Science and Technology (CiSt)*, 2020, pp. 1–6.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details

www.ijircce.com