

# A Comprehensive Review of Machine Learning and Explainable AI Approaches for Road Accident Severity Prediction

SHIKHA PATEL<sup>1</sup>, DR. D. GANESH<sup>2</sup>

<sup>1</sup>Scholar (MSC CS &IT), Department of Computer Science & IT, Jain (Deemed-To-Be-University) Bangalore, India

<sup>2</sup>Professor, Department of Computer Science & IT, Jain (Deemed-To-Be-University) Bangalore, India

*Abstract- Road traffic accidents continue to be one of the major causes of injuries and fatalities across the globe, leading to severe social and economic consequences. In recent years, advancements in artificial intelligence have encouraged the use of machine learning (ML) and deep learning techniques to analyze accident data and predict accident severity. A variety of predictive approaches, including Random Forest, Neural Networks, Support Vector Machines, and deep learning architectures, have been explored to improve accident severity classification. Despite these developments, many existing studies face challenges such as limited dataset sizes, poor model interpretability, class imbalance problems, and lack of real-time implementation. Moreover, the transparency of predictive models has become an essential requirement for their practical use in transportation safety systems. This study provides a comprehensive review of fifteen recent research works focusing on accident severity prediction, hotspot identification, driver behavior analysis, and explainable artificial intelligence (XAI). The research compares different methodologies, datasets, and performance outcomes in order to highlight current research trends and identify gaps in the literature. Based on this analysis, a conceptual framework integrating machine learning, deep learning, and explainable AI techniques is proposed to enhance prediction accuracy and model interpretability. The proposed approach aims to support intelligent transportation systems and assist decision-makers in improving road safety strategies.*

**Keywords - Machine Learning, Accident Severity Prediction, Explainable Artificial Intelligence, Deep Learning, Road Safety, Random Forest**

## I. INTRODUCTION

### 1.1) Background of the Study

Road traffic accidents are a major global threat to the overall safety of people with over 1.3 million

individuals dying annually as a result of accidents. The high population expansion in cities, growth in the quantity of vehicles on the road, and unlawful driving are major factors that lead to the occurrence of road accidents. The ability to forecast the severity of accidents is valuable to enhance emergency services, design safer roads, and come up with smart transportation systems. Machine learning methods have been useful in the examination of massive traffic data and the identification of trends that dictate the severity of accidents. Some of the algorithms used to enhance prediction performance and improve the decision-making abilities include the Random Forest, the Neural Networks, the Support Vector Machines, and the Gradient Boosting algorithms.

### 1.2) Problem Statement

Despite a number of studies devoted to the severity of accidents prediction, there are still numerous challenges: Little explanation of multifaceted machine learning models. Small or unbalanced data used. Absence of systems that predict accidents on time. The lack of explainable artificial intelligence approaches. Such constraints decrease the predictive model effectiveness and limit its use in contemporary transport systems.

### 1.3) Motivation

Proper forecasting of the severity of accidents can be a key in improving the road safety. It may guide emergency services in quick response, assist authorities in enhancement of traffic control sites and cooperation among policymakers in making wise decisions. Due to the development of machine learning and clarifiable AI, it is now possible to create intelligent systems that are precise and clear.

#### 1.4) Objectives of the Study

The principal aims of this study are: To review the current machine learning and deep learning methods in prediction of accidents severity. To review and compare different models, data sets and methodologies used in the past. To detect research gaps in the existing systems of accident prediction. To suggest a conceptual framework that would combine machine learning and explainable AI methods.

#### 1.5) Contributions of the Paper

This research contributes the following:

- A detailed review of fifteen recent research studies related to accident prediction
- Comparative evaluation of different machine learning techniques
- Identification of important research gaps in existing literature
- Proposal of a hybrid framework combining machine learning and explainable AI

#### 1.6) Organization of the Paper

The remainder of this paper is organized as follows:

Section 4 presents the Introduction

Section 5 presents the literature review and comparative analysis of existing research.

Section 6 discusses the identified research gaps.

Section 7 explains the proposed methodology.

Section 8 presents expected results and potential applications.

Section 9 concludes the paper.

## II. LITERATURE REVIEW

### 2.1) Thematic Classification of Literature

Research related to accident severity prediction has progressed rapidly with the development of machine learning and artificial intelligence technologies. Existing studies can generally be categorized into four groups: machine learning approaches, deep learning approaches, real-time prediction systems, and explainable AI-based methods.

#### 2.1.1) Machine Learning Approaches

A number of researchers have applied machine learning algorithms to the historical data on accidents and forecasting the severity of accidents.

The present research paper by Ahmed et al. (2021) involves a comparative study of the various machine learning algorithms namely Random Forest, Logistic Regression, K-Nearest Neighbors, Naive Bayes, XGBoost, and AdaBoost. In their paper, they stated that the Rand Forest model gave the highest performance with a prediction accuracy of 86.64% in classifying the severity of accidents as binary.

Gupta et al. (2022) proposed a system that uses the Kernel Density Estimation with machine learning models to determine the areas that are prone to accidents and estimate the level of severity of accidents. Their study has proved that predictive modeling can be used to align the spatial analysis with accuracy in the assessment of hazards that involve accidents.

Katoch et al. (2023) compared the use of various classification algorithms; KNN, Support Vector Machines, and Neural Networks with accident data gathered in India. Their findings suggested that neural network models achieved the largest prediction accuracy that was about 90 percent.

#### 2.1.2) Deep Learning Approaches

The forecasting of the intensity of the accidents has been experiencing increasing attention due to the method of deep learning that learns complex patterns using a large set of data. A hybrid model has also been proposed by Rahul Kumar et al. (2025) and incorporates Convolutional Neural Networks and machine learning algorithms such as XGBoost and Random Forest. Their hypothesis experiment showed that CNN-XGBoost model was accurate by 95%. Similarly, in their design of an Artificial Neural Network, Jameel Sowdagur et al. (2020) created a Multi-Layer Perceptron model that predicted the level of accidents. The proposed system set an accuracy of 84.1% following the optimization of the tuning of the model parameters.

#### 2.2) Comparative Analysis of Existing Methods

Paper	Method	Dataset	Accuracy	Limitations
Ahmed et al. (2021)	RF, LR, KNN	NZ data	86.64%	Limited human factors
Gupta et al. (2022)	KDE + ML	UK data	High accuracy	Static spatial modeling
Kumar et al. (2025)	CNN + XGBoost	Traffic dataset	95%	High computational cost
Sowdagur et al. (2020)	ANN	Mauritius dataset	84.1%	Manual tuning
Hussain et al. (2025)	RF, DT	Incident data	98.45%	Overfitting risk

### 2.3) Critical Review

However, despite the fact that research studies have shown stunning outcomes in forecasting accuracy of accidents, several issues have been experienced. The association of complexities that are represented in a large scale accidents dataset are usually weak to older machine learning models. Despite the fact that deep learning methods help to improve the outcome of prediction, it is also linked to the extensive usage of computational resources and enormous training data. In addition, the majority of the available publications have adopted geographically constrained datasets that limit the capability of the models to generalize to other geographical territories. The other notable limitation is that most predictive models cannot be understood and, therefore, a transportation authority cannot easily understand what the prediction is founded on. The recent past has also witnessed clear artificial intelligence approaches to solve these issues by increasing the level of transparency and providing insights into explainability in prediction models.

### 2.4) Identified Research Gaps

As it turned out, there are several gaps in research associated with the literature review: There is slight application of explainable AI, in models which predict the severity of accidents. Dependence on small datasets or datasets that are restricted by geographical means. Lack of hybrid systems, which integrate deep-learning and clarifiable AI. Weak development of real time accident predicting systems.

## III. PROPOSED METHODOLOGY

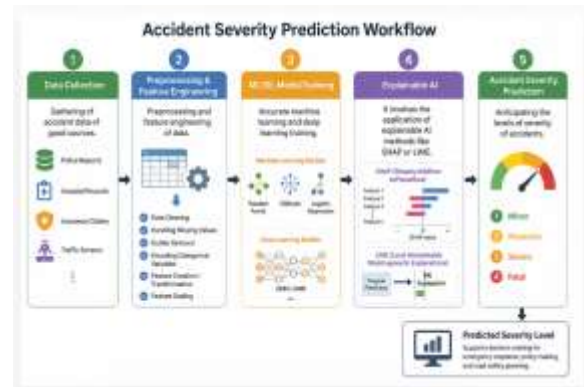
### 3.1) System Overview

The proposed framework includes machine, deep and explainable AI approach to improve the quality and interpretability of the accident severity prediction systems

### 3.2) Workflow

The suggested system adheres to the subsequent working process:

- Gathering of an accident data of good sources.
- Preprocessing and feature engineering of data.
- Accurate machine learning and deep learning training.
- It involves the application of explainable AI methods like SHAP or LIME.
- Anticipating the levels of severity of accidents.



### 3.3) Dataset Description

The publicly available accident datasets that can be used in the proposed system include:

- US Accident Dataset
- UK Traffic Accident Dataset
- Indian Road Accident Dataset

These datasets contain important variables including weather conditions, road characteristics, vehicle speed, time of accident, and driver behavior.

## IV. EXPECTED RESULTS AND DISCUSSION

### 4.1) Expected Outcomes

The suggested framework should result in the following results:

- Better accuracy of prediction than traditional machine learning models.
- Increased explainability in prediction outcomes with explainable AI techniques.
- Ability to aid in real time risk prediction of accidents.

#### 4.2) Comparative Evaluation

The suggested structure will be contrasted with the currently available algorithms such as:

- Random Forest
- Neural Networks
- XGBoost
- Support Vector Machines

Performance will be evaluated using standard evaluation metrics such as accuracy, precision, recall, and F1-score.

#### 4.3) Discussion

Accomplishing machine learning with explainable AI methods can be highly beneficial to accident severity prediction systems. Conventional machine learning algorithms can be highly accurate in their predictions, although they can be black-box in nature, and thus prediction results cannot be easily interpreted.

The proposed framework will be able to produce explainable prediction results by including explainable AI methods including SHAP and LIME. It is this openness that is vital towards gaining trust between the transportation authorities, safety analysts, and policymakers.

Moreover, use of advanced machine learning algorithms used to analyze big sets of accidents may assist in defining key aspects affecting the severity of accidents, such as weather conditions, road infrastructure, traffic density, and driver behavior. Learning about these factors will be useful in the context of designing successful road safety policies

### V. APPLICATIONS AND USE CASES

The suggested framework can be utilized in a variety of spheres such as:

- Smart transportation systems.
- Traffic monitoring of a smart city.
- Management of emergency response.
- Canada: driver safety assistance technology

### VI. CONCLUSION

The present paper was a thorough overview of machine learning and deep learning algorithms that are used in predicting the severity of accidents. The fifteen recent studies have been analyzed, and it was possible to point out the great improvement of the prediction performance with the help of such algorithms as Random Forest, Neural Networks, and CNN-based models.

Nevertheless, the current systems are still associated with the issues of model interpretability, limitations of the datasets, and the absence of real-time implementation. As a solution to these difficulties, this paper presented a conceptual framework of the combination of machine learning and explainable AI methods.

The future studies need to work on the design of scalable, interpretable, and real-time accident prediction systems that can be used to improve road safety and facilitate intelligent transportation infrastructure.

### REFERENCES

- [1] S. Ahmed et al., "Comparative Study of Machine Learning Algorithms for Accident Severity Prediction," IEEE IUCC, 2021.
- [2] U. Gupta et al., "Hotspot Analysis and Severity Prediction Using Machine Learning," IEEE Conference, 2022.
- [3] R. Kumar et al., "Deep Learning-Driven Accident Severity Prediction," IEEE, 2025.
- [4] V. Datta et al., "ML-Powered Web Application for Accident Severity Prediction," Conference Proceedings, 2023.

- [5] D. Dinesh, "Real-Time Accident Prediction System Using Deep Neural Networks," Conference Proceedings, 2023.
- [6] R. Shafique et al., "AV Sensor-Related Accident Severity Prediction Using Ensemble Learning," IEEE Access, 2024.
- [7] J. A. Sowdagur et al., "Artificial Neural Network for Accident Severity Prediction," Conference Proceedings, 2020.
- [8] M. Evin et al., "Alcohol Functional State Detection Using Machine Learning," IEEE BIBM, 2018.
- [9] K. Kim and Y. Lim, "Interpretable Maritime Accident Prediction Using Explainable AI," IEEE Access, 2022.
- [10] N. Kılıç et al., "Driver Injury Severity Prediction Using SMOTE-ENN and Random Forest," Conference Proceedings, 2023.
- [11] M. Mishra et al., "Spatio-Temporal Accident Count Prediction Using Machine Learning," Conference Proceedings, 2024.
- [12] K. Hussain et al., "Passenger Safety Enhancement Using Machine Learning Techniques," IEEE SENNET, 2025.
- [13] R. Mynavathi et al., "Machine Learning-Based Accident Severity Prediction with Explainable AI," Conference Proceedings, 2023.
- [14] K. Katoch et al., "Machine Learning-Based Classification of Road Accident Severity in India," Conference Proceedings, 2023.
- [15] N. Srivastava et al., "Unraveling Road Accident Risk Prediction Models With Explainable AI," IEEE Transactions on Computational Social Systems, 2025.
- [16] WHO, "Global Status Report on Road Safety," World Health Organization, 2023.
- [17] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.
- [18] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
- [19] C. Cortes and V. Vapnik, "Support-Vector Networks," Machine Learning, vol. 20, pp. 273–297, 1995.
- [20] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," Advances in Neural Information Processing Systems, 2017.
- [21] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You? Explaining the Predictions of Any Classifier," KDD Conference, 2016.
- [22] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," KDD Conference, 2016.
- [23] D. Silver et al., "Mastering the Game of Go with Deep Neural Networks and Tree Search," Nature, 2016.
- [24] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.
- [25] J. Schmidhuber, "Deep Learning in Neural Networks: An Overview," Neural Networks, vol. 61, pp. 85–117, 2015.