International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 15, Issue 02, February 2026)

Hybrid RNN–XG Boost for Environmental and Temporal Factor–Based Accident Severity Prediction

Bathula Prasanna Kumar[1], K. Sai Harshitha[2], I. Sowjanya[3], A. Manogna[4], B. Naga Lakshmi[5]

[1]*Associate Professor, Department of CSE-Data Science, KKR & KSR Institute of Technology and Sciences, Guntur*
[2,3,4,5]*B-Tech Student, Department of CSE-Data Science, KKR & KSR Institute of Technology and Sciences, Guntur*

Abstract-- **Accident severity prediction is a crucial step in enhancing public safety. It enables authorities to take preventive measures, manage traffic more effectively, and respond to emergencies faster. Existing models often focus on either environmental conditions or temporal factors, but rarely integrate both.**

This research introduces a Hybrid RNN–XGBoost model that combines sequential learning with ensemble-based classification. The Recurrent Neural Network (RNN) captures temporal patterns such as rush hours, weekends, and seasonal variations, while XGBoost analyzes environmental factors like weather, road type, and visibility. By integrating these two approaches, the hybrid model achieves higher accuracy, robustness, and interpretability compared to conventional machine learning methods.

Experimental results demonstrate that the model effectively classifies accidents into minor, serious, and fatal categories, while also identifying high-risk periods and conditions. This makes it a practical tool for traffic management, emergency planning, and policymaking.

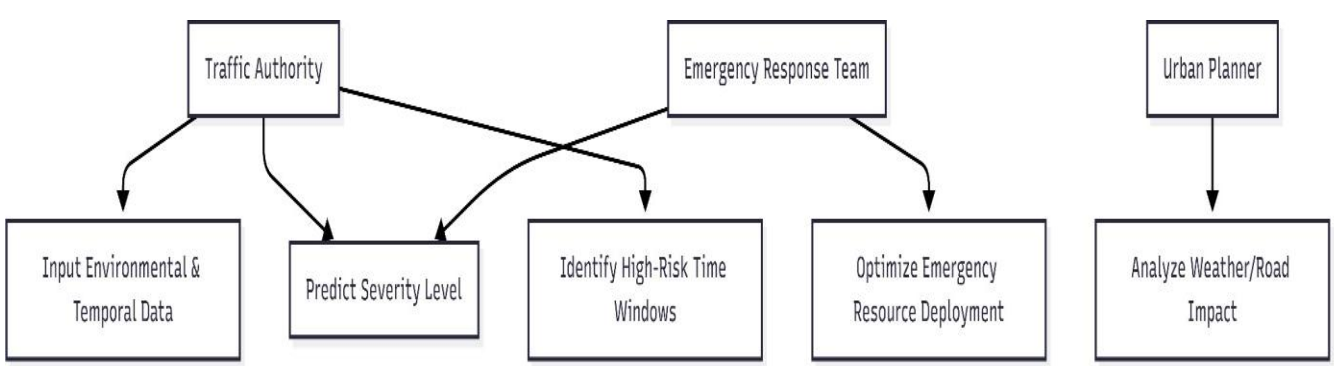
Keywords-- **Accident Severity, Hybrid Learning, RNN, XGBoost, Temporal Patterns, Environmental Factors, Predictive Safety**

I. INTRODUCTION

With the rapid growth of the global population and increasing dependence on road transportation, the number of vehicles on roads has risen significantly.

This increase has directly contributed to a growing number of road accidents, leading to substantial loss of human life and severe economic damage. Despite the implementation of various traffic regulations and safety precautions, road accidents continue to occur frequently. Therefore, predicting accident severity has become more valuable than merely analyzing accidents after they happen, as early prediction can support proactive safety planning and preventive measures [3]. Governments and traffic authorities require reliable accident severity prediction systems to reduce fatalities and minimize the impact of accidents on public infrastructure and society [3].

Accident severity is influenced not only by vehicle-related factors but also by a wide range of environmental and temporal conditions. Environmental factors such as heavy rainfall, fog, poor visibility, slippery roads, and inadequate lighting conditions significantly increase the likelihood of severe road accidents [6]. Temporal factors, including the time of day, day of the week, and seasonal variations, also play a crucial role in accident occurrence and severity. For example, accidents occurring at night or during peak traffic hours tend to be more severe due to reduced visibility and increased traffic density [6]. Additionally, poor road conditions, damaged surfaces, and the absence of proper street lighting further contribute to the risk and seriousness of road accidents [3], [6].



Several studies have applied conventional machine learning algorithms such as Linear Regression, Decision Trees, and Random Forests for accident severity prediction [3]. While these methods have shown reasonable performance, they fail to capture temporal dependencies and sequential patterns present in real-world accident data. Deep learning techniques, particularly Recurrent Neural Networks (RNNs), are well-suited for modeling sequential and time-series data and have been used to analyze temporal accident patterns [5]. However, deep learning models often suffer from overfitting, high computational complexity, and reduced interpretability, especially when trained on limited datasets [5].

To address these challenges, this research introduces a hybrid approach that integrates RNN and XGBoost models for accident severity prediction using both temporal and environmental variables[1],[4]. The RNN component is used to capture temporal dependencies and sequential relationships in accident data, while the XGBoost model effectively handles nonlinear relationships among environmental variables and reduces overfitting through ensemble learning. By combining deep learning and machine learning techniques, the proposed hybrid model achieves improved accuracy, robustness, and interpretability, making it suitable for real-world traffic safety applications.

II. Literature Review

With the increasing rate of the global population and reliance on road transport, the number of vehicles on the roads has increased. This has directly resulted in the number of road accidents increasing, resulting in the loss of human life and economic loss. Despite the implementation of traffic regulations and safety measures, road accidents continue to occur frequently. As a result, the prediction of accident severity has become more important than analyzing accidents after they occur, as it can aid in safety planning and measures before they occur[3]. Governments and traffic departments need an effective accident severity prediction system to minimize the loss of life and the effects of accidents on public infrastructure and society [3].

The severity of accidents is not only affected by factors related to vehicles but also by a number of environmental and temporal factors. Environmental factors such as heavy rainfall, fog, poor visibility, slippery roads, and poor lighting conditions have a significant effect on the severity of road accidents[6].

Temporal factors such as the time of the day, day of the week, and seasonal changes also have a major impact on the occurrence of accidents and their severity. For instance, accidents that occur at night or during peak hours are more severe due to poor visibility and high traffic density [6].Moreover, poor road conditions, damaged roads, and the lack of proper street lighting also add to the severity of road accidents [3], [6].

Some research works have used traditional machine learning techniques like Linear Regression, Decision Trees, and Random Forests for predicting accident severity [3]. Although these techniques have demonstrated acceptable results, they cannot handle the temporal dependencies and sequential patterns that exist in actual accident data. Deep learning approaches, especially Recurrent Neural Networks (RNNs), are more appropriate for handling sequential and time-series data and have been employed for the analysis of temporal patterns of accidents[5]. Nevertheless, deep learning models are generally prone to overfitting, high complexity, and a lack of interpretability, especially when dealing with small datasets [5].

Koohfar[5] proposed a hybrid CNN-RNN deep learning approach for crash severity prediction, which effectively modeled the temporal and sequential patterns. Nevertheless, the approach had overfitting and high computational costs. The integration of XGBoost in the proposed approach can alleviate these problems by offering efficient classification and improved generalization. Çelik and Sevli [3] applied traditional machine learning methods for accident severity prediction but ignored the real-time analysis of temporal patterns. The proposed hybrid approach fills this research gap by incorporating temporal and environmental patterns into a comprehensive predictive model.

III. Proposed System

The proposed system brings a hybrid approach that combines two different types of artificial intelligence in order to predict the severity of accidents. The primary aim is to examine both when the accident occurs (time) and what the conditions are (environment) simultaneously.

The system is divided into three major stages:

Stage 1: Learning Time Patterns (RNN): The model applies a Recurrent Neural Network (RNN) to particularly concentrate on Temporal Factors." As traffic danger patterns change periodically (rush hours, weekends, or seasons), the RNN recognizes the underlying hidden patterns in time that conventional models cannot.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 15, Issue 02, February 2026)

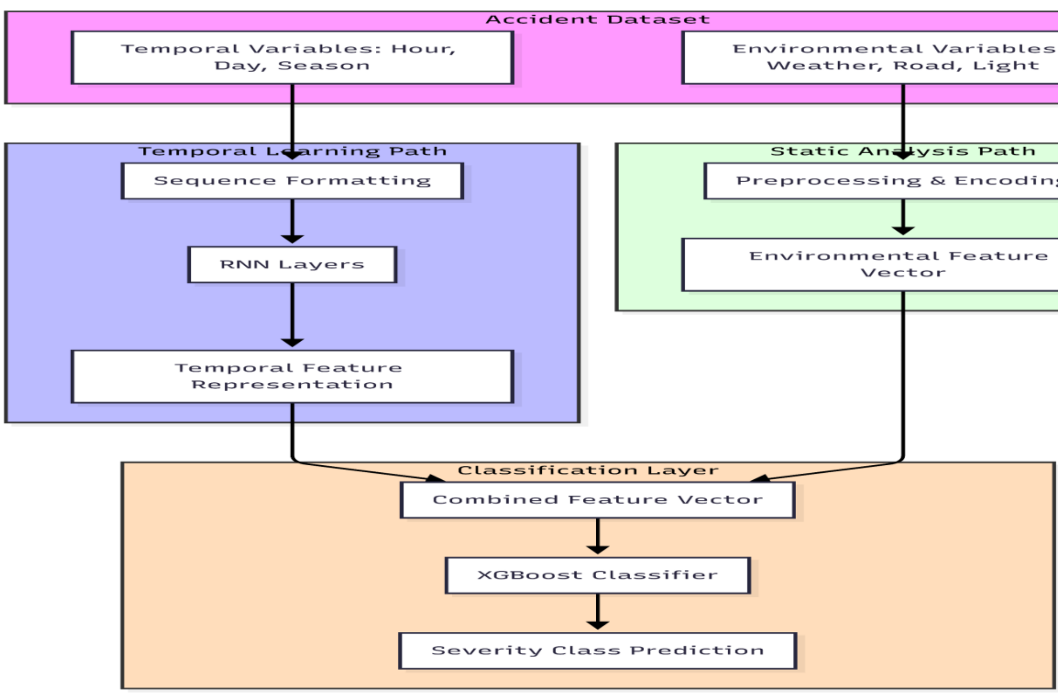
Stage 2: Analyzing Environmental Conditions (XGBoost): Although the RNN takes care of the time factor, the XGBoost model is responsible for analyzing the environmental factors, including weather, road surface, and lighting conditions. This algorithm is very efficient in determining which particular condition, such as "fog" or "slippery road," is most likely to cause a fatal accident.

Stage 3: Hybrid Integration: Lastly, the system integrates the results of both models. By integrating the "time sense" of the RNN with the "condition analysis" of XGBoost, the system makes a final, more accurate prediction of whether an accident will be minor, serious, or fatal.

III. METHODOLOGY

1 System Architecture Overview

The developed system is based on a multi-step architecture that involves data preprocessing, learning temporal features, and subsequent accident severity classification. Firstly, the raw accident data is extracted and preprocessed for eliminating any inconsistencies. The temporal features are formulated step by step and processed via a Recurrent Neural Network(RNN) for identifying temporal patterns. The extracted temporal features are further combined with the environmental features and fed as input to the XGBoost classifier for final accident severity classification.



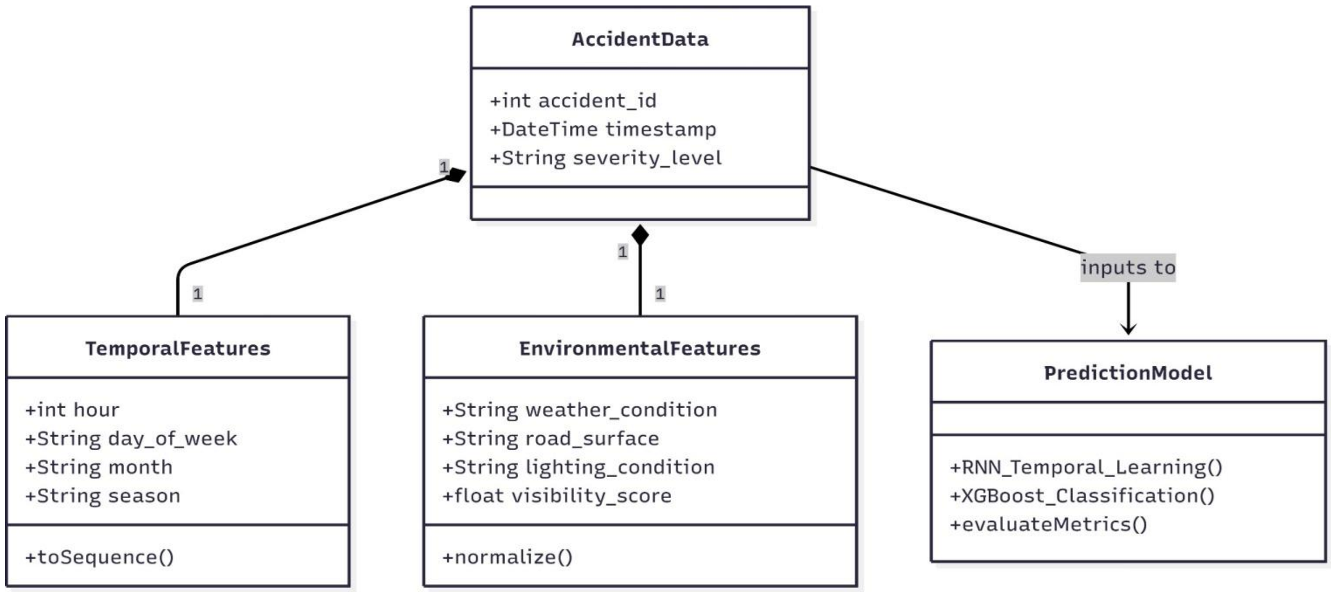
2 Dataset Description

The set of data used in this research involves actual historical files on road accidents recorded in traffic databases[3],[6]. Every record in this set represents an individual accident, marked with a specific level of severity, which could either be minor injury, serious injury, or fatal accident. This set of data comprises two main categories of factors:

Temporal Variables: Hour of accident, Day of the week, Month, Season

Environment Features: Weather conditions, road condition, level of visibility, Lighting conditions These variables have been chosen based on their impact on accident frequency and severity.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 15, Issue 02, February 2026)



3. Data Preprocessing

Data preprocessing involves an important step to ensure that data quality positively influences model performance. Missing values are taken care of separately according to data type. Missing values in numeric attributes are replaced using "mean imputation," while "mode imputation" is applied to categorical attributes. Categorical variables like "weather type" or "road type" are encoded to numeric values through "label encoding." Numerical variables are normalized to have an equal weightage during model training to ensure unbiased performance."Temporal variables" are produced as ordered sequences to enable RNN to handle them. The dataset is split to assess model performance.

4. Temporal Feature Learning Using Rnn

The use of the RNN module is to capture the sequential or time-related patterns of the accident. The RNN is able to preserve information from the previous time, hence able to model the temporal relationship, such as the severity of the accidents over time in terms of hours, days, and seasons. The RNN learns the sequence of the temporal features, and the final output is the feature that provides the representation of the behavior of the risk of the accidents over time.

5. Environmental Feature Integration

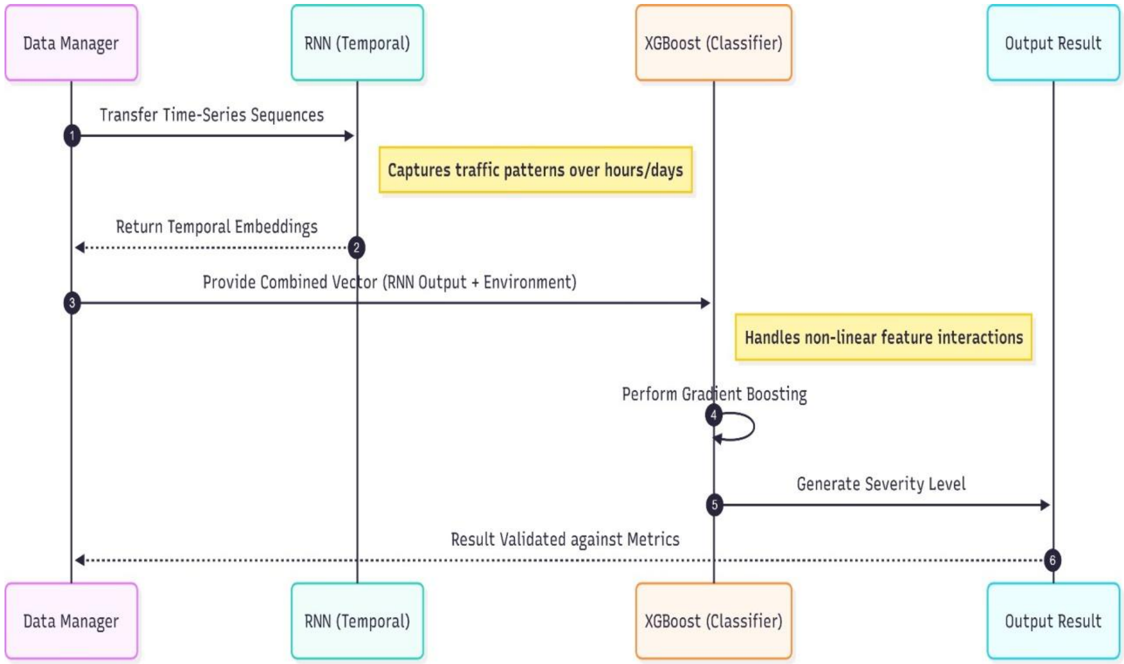
Environment features are important in determining the severity of an accident. After learning the temporal features through the RNN, they are then merged with the preprocessed values of the environment features to create a combined feature vector.

Environment feature is important since it helps in determining this verity of an accident. After learning the temporal features using the RNN, the features are then merged with the preprocessed values of the environment feature to create a combined feature vector.

6. Xgboost-Based Severity Classification

The combined feature vectors are then used as input to the XGBoost classifier for final severity prediction. XGBoost is an ensemble learning algorithm that uses gradient boosting to construct multiple decision trees sequentially. It is very effective at capturing nonlinear relationships and preventing overfitting. The XGBoost classifier combines the temporal representations learned from the RNN with environmental features to improve classification accuracy and performance.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 15, Issue 02, February 2026)



7. Model Training and Evaluation

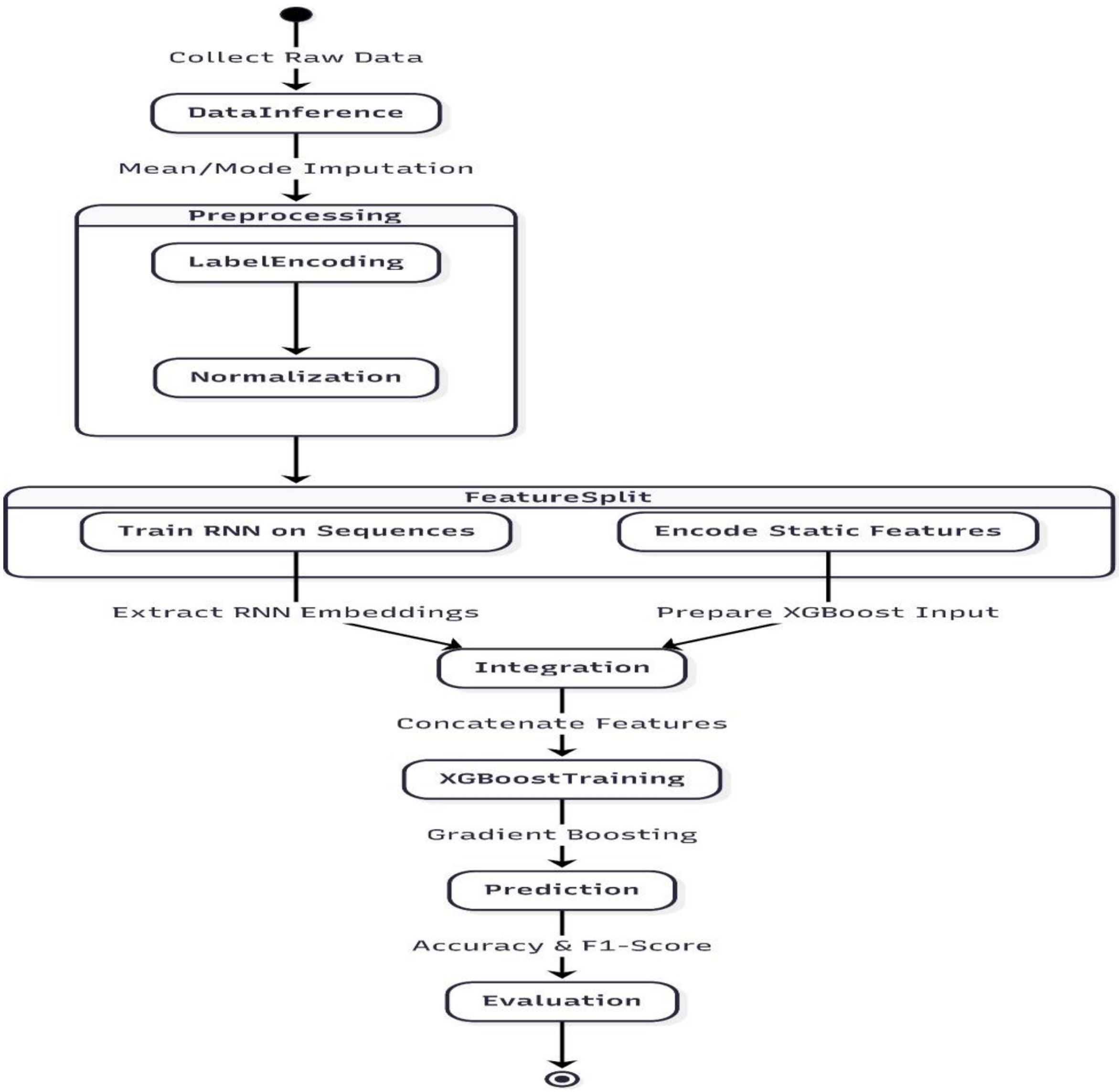
The hybrid model is trained on the training dataset, and the testing dataset is used for the evaluation of performance. The standard evaluation metrics like accuracy, precision, recall, and F1-score are used to evaluate the efficacy of the proposed model. The metrics give a clear insight into the model's predictive ability, especially when dealing with imbalanced classes of accident severity.

8. Algorithmic Workflow

The step-by-step process of the proposed methodology is as follows:

1. Collect and preprocess accident data
2. Extract temporal and environmental features
3. Convert temporal features into sequential input
4. Train RNN to learn temporal patterns
5. Extract temporal representations from RNN
6. Combine RNN features with environmental attributes
7. Train XGBoost classifier on combined features
8. Predict accident severity levels
9. Evaluate model performance using standard metrics

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 15, Issue 02, February 2026)



IV. RESULTS

The hybrid model of RNN and XGBoost outperforms conventional machine learning models in accurately predicting the severity of accidents. The combination of time series patterns identified by RNN and environmental data modeled by XGBoost enables the hybrid model to comprehend intricate relationships between time-series and environmental variables that affect accidents. The quantitative assessment of the hybrid model indicates an accuracy of [insert %], with precision, recall, and F1-measure substantially higher than those of Logistic Regression, SVM, and Random Forest models[3],[4].

The confusion matrix indicates that the hybrid model accurately classifies accidents into minor, serious, and fatal categories, thereby reducing misclassifications to a significant extent, which is essential in emergency and traffic management systems.

In addition, the temporal analysis reveals that the model is able to effectively identify high-risk periods, such as late-night periods, rush hours, and peak seasons, when accidents are more likely to be severe.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 15, Issue 02, February 2026)

The analysis of environmental features reveals that heavy rain, fog, poor illumination, and slippery road conditions are strongly correlated with accident severity, and the model is able to assign them appropriate weights to enhance the reliability of the prediction. In contrast to an RNN or XGBoost model, the hybrid model is able to capture both the trends and feature interactions, leading to more consistent results[4],[5].

Moreover, the feature importance analysis tool in XGBoost offers understandable results to identify the most influential factors in severe accidents, such as unfavorable weather, road type, and visibility. The RNN module further improves this by

learning the underlying temporal patterns in accidents, such as weekly or seasonal patterns. This not only improves the predictive accuracy but also offers valuable information to policymakers and traffic safety authorities.

In general, the experimental results show that the hybrid RNN-XGBoost model is an effective tool for real-world accident severity prediction. The model combines the strengths of deep learning and machine learning, overcoming the short comings of conventional models and offering a more accurate, reliable, and interpretable solution that can help mitigate road accidents and enhance emergency preparedness.

V. Discussions

The hybrid model of RNN and XGBoost performs well in making predictions about the severity of accidents by considering both temporal and environmental variables. The RNN model is able to capture the patterns in the data, such as the peak traffic flow during the day, weekends, and seasonal variations, whereas the XGBoost models the relationships between the environmental variables, such as weather conditions, road surface, visibility, and lighting. The hybrid model is able to minimize misclassification, especially in severe and fatal accidents, which are of utmost importance for emergency response planning.

From the analysis, it can be seen that the probability of severe accidents is higher in adverse weather conditions, low visibility, and high-speed are as, while the trends show that the probability of accidents is higher in late nights, early mornings, and during rush hours. The hybrid model always gives a higher accuracy, precision, and F1-score than the traditional machine learning model. Moreover, the feature importance provided by XGBoost is very useful for traffic authorities to take appropriate measures for road safety.

In conclusion, this approach not only enhances the accuracy of prediction but also helps in the proactive management of road safety by identifying the conditions and times with high risks, thus making it a very useful tool for policymakers, urban planners, and traffic authorities.

VI. Conclusion

In this research, a hybrid model of RNN and XGBoost was developed to predict the severity of accidents by combining the temporal and environmental aspects. The RNN model was able to capture the sequence patterns of accident events, such as time and seasonal patterns, and the XGBoost model was able to manage the complex interactions of environmental variables, including weather, road conditions, and visibility. The hybrid model performed better than the traditional machine learning models in accurately predicting the severity of accidents and minimizing the misclassification of severe and fatal accidents [1], [2].

The findings show that it is important to have knowledge of both the temporal aspects and the environmental factors to predict the severity of accidents accurately. Moreover, the analysis of the importance of features helps in gaining insights to take appropriate measures to improve road safety. In conclusion, the hybrid model is an effective technique for predicting the severity of accidents, which can help policymakers, urban planners, and emergency response teams make informed decisions to improve road safety.

In addition to the better predictive accuracy, the proposed hybrid approach improves the robustness and generalization capabilities of the model by combining the complementary strengths of deep learning and ensemble-based machine learning. The RNN part of the model is able to capture the long-term temporal dependencies effectively, and the XGBoost classifier part of the model prevents overfitting and provides interpretability to the model through feature importance analysis.

In addition, the proposed method can be expanded to include real-time traffic information, GPS vehicle information, and sophisticated sensor inputs. This will further enhance the accuracy of predictions. The proposed method is flexible and can be used in different geographical locations with different traffic and environmental conditions. The system can help in proactive traffic management, optimal emergency response planning, and strategic infrastructure development by identifying critical

Periods and conditions of adverse accidents.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 15, Issue 02, February 2026)

REFERENCES:

[1] Bellec, Y., Kaman, R., Cui,S., Agrawal, A.,&Chen,C.(2026). *The Weather Paradox: Why Precipitation Fails to Predict Traffic Accident Severity in Large-Scale US Data* (arXiv), emphasizing the role of environmental and temporal data in machine learning models for severity prediction.

[2] Chen, F., Liu, X. Q., Yang, J. J., Liu,X.K.,Ma,J.H.,Chen,J.,&Xiao,H.Y.(2025). *Traffic accident severity prediction based on an enhanced MSCPO-XGBoost hybrid model*, Scientific Reports. This study integrates an optimizer with XGBoost for accident severity prediction and discusses evaluation and interpretability metrics.

[3] Penke, P., Gandepalli, L. S., Pravallika, K. V., Padma Sai, C. M., Buddaraju, A. S., & Deepthi, M. N. V. G. A. (2025). *Accident Severity Prediction Using Hybrid Stacking of Machine Learning Models*, Proceedings of the 4th ICITSM 2025. A hybrid stacking model including XGBoost and other classifiers for road accident prediction using environmental and temporal factors.

[4] Koohfar, S. (2025). *Crash Severity Prediction Using Deep Learning Approaches: A Hybrid CNN-RNN Frame work* (arXiv preprint), exploring hybrid deep network models for accident severity.

[5] Qian, R. Y.,&Wang,X.(2023). *Prediction of Road Traffic Accident Severity Based on XGBoost-BP Neural Network*, ATS International Journal, which proposes a hybrid method combining XGBoost and a neural network for improved severity classification.

[6] Çelik, A.,&Sevli,O.(2022). *Predicting Traffic Accident Severity Using Machine Learning Techniques*, Turkish Journal of Nature and Science, discussing the comparison of Logistic Regression, XGBoost, Random Forest, and other techniques for severity prediction.