Quantum Pulse:
Journal of AI, Data Science & Systems Engineering
ISSN: 3108-1061 | Volume 2 Issue 1 January – June 2026 | QP15

EFFICIENT ROAD ACCIDENT SEVERITY CLASSIFICATION USING RF-RFE AND DEEP LEARNING FRAMEWORK

[#1]**KOLIPAKA MADHUSRI,** *Dept of CSE,*
[#2]**Dr.D.SRIKANTH REDDY,** *Assistant Professor, Dept of CSE,*
Vaageswari College of Engineering(Autonomous), Karimnagar, TG.

ABSTRACT: The purpose of this research is to enhance the precision and dependability of traffic accident severity predictions by introducing a robust framework for classifying road accident severity that incorporates Random Forest–Recursive Feature Elimination (RF-RFE) and a deep learning model. This work uses RF-RFE for optimal feature selection to reduce dimensionality and eliminate unnecessary variables from large accident datasets, with the goal of identifying the most important factors. The selected criteria are then fed into a deep learning system that can detect complex nonlinear relationships between variables including road conditions, weather, vehicle specs, and driver behaviors. Through the integration of deep neural network classification and machine learning-driven feature optimization, the suggested technique improves prediction accuracy, streamlines computations, and enables real-time decision-making. In terms of data categorization and resilience, the results show that the hybrid RF-RFE and deep learning methodology outperforms standard models. For this reason, it is a useful tool for traffic management authorities and lawmakers to use in their pursuit of safer roads and less severe accidents.
Keywords: *Road Accident Severity, Random Forest, Recursive Feature Elimination (RF-RFE), Deep Learning, Feature Selection, Classification Model, Machine Learning.*

1. INTRODUCTION

Road traffic accidents constitute a significant global public safety concern, resulting in severe injuries, numerous fatalities, and substantial financial costs for remediation. Accelerated urbanization, an increase in vehicular traffic, and expanded transit networks have rendered traffic systems more complex. This complicates the management and prevention of accidents. The prediction and classification of traffic accident severity is an increasingly significant research domain, as it can assist emergency responders, legislators, and transportation authorities in making prompt and educated decisions. Advanced data-driven techniques are increasingly employed to analyze extensive accident data to uncover concealed trends that influence injury outcomes.

Road crashes can be categorized into numerous levels, including minor injuries, catastrophic injuries, and fatal accidents. These levels are contingent upon several things. Factors include the driver's proficiency, the vehicle type, the accident vicinity, road design, time of day, and traffic volume.

Conventional statistical models often struggle to characterize the intricate nonlinear relationships among these variables, particularly in high-dimensional and unbalanced

Paper Available at: www.qpjournal.org
DOI : https://doi.org/10.65713/qpv2i115

datasets. Machine learning techniques are increasingly popular due to their ability to process vast datasets and predict complex interactions among many characteristics.

Feature selection is essential for enhancing the efficacy and performance of accident severity classification algorithms. In high-dimensional data, superfluous or unnecessary features can diminish model accuracy and complicate processing. The Random Forest–Recursive Feature Elimination (RF-RFE) method is an efficient approach for identifying the most significant predictors. This is accomplished by systematically eliminating less significant characteristics according to relevance ratings derived from Random Forest models. RF-RFE improves the interpretability of the predictive framework, reduces overfitting, and optimizes classification performance by identifying an optimal set of pertinent features.

Deep learning techniques enhance the prediction of accident severity by identifying intricate nonlinear patterns and hierarchical representations within the data. Deep feedforward networks and convolutional neural networks are two neural network architectures capable of autonomously identifying significant feature representations in both organized and unstructured data. When employed alongside RF-RFE-based feature selection, deep learning models can concentrate on the most significant variables. This enhances their efficiency and improves their generalization capabilities. This hybrid approach integrates the optimal elements of deep neural networks and ensemble learning to ensure precise severity categorization.

The suggested efficient framework for classifying the severity of traffic accidents incorporates a deep learning model for precise prediction alongside RF-RFE for optimal feature selection. The objective of this integrated technique is to expedite processing, simplify models, and enhance classification accuracy. The framework offers a scalable and dependable solution for intelligent transportation systems by utilizing optimized features and sophisticated learning algorithms. This facilitates data-driven approaches for risk assessment, accident prevention, and resource allocation in contemporary traffic management systems.

2. EFFICIENT SEVERITY CLASSIFICATION FRAMEWORK

Data Description

The dataset for this research was sourced from the official French national road accident database, accessible to the public via government traffic safety records. This information is derived from comprehensive police records and traffic accident reports. The dataset has four files: Characteristics, Places, Users, and Vehicles. Each file contains extensive information regarding the vehicle, the roadway infrastructure, the individuals utilizing the route, and the circumstances surrounding the event.

The objective variable is accident severity, categorized into four types.

- Uninjured
- Minor injury
- Serious injury
- Fatal injury

The severity of accidents varies significantly between classes. Fatal accidents are exceedingly uncommon in comparison to minor and non-injurious incidents.

Paper Available at: www.qpjournal.org
DOI : https://doi.org/10.65713/qpv2i115

Quantum Pulse:
Journal of AI, Data Science & Systems Engineering
ISSN: 3108-1061 | Volume 2 Issue 1 January – June 2026 | QP15

File Name	Feature Name	Number of Unique Values (If Categorical)
Characteristics	Month	12
	Day	7
	Hour	Time
	Luminosity	5
	Agglomeration	2
	Intersection	8
	Weather conditions	9
	Collision type	8
	Geographic coordinates: Latitude, Longitude	Number
Places	Road category	8
	Traffic direction	5
	Number of lanes	7
	Road gradient	5
	Road alignment	3
	Road width	Number
	Surface condition	10
	Infrastructure	10
	Accident location	8
	Maximum authorized speed	Number
Users	Occupied place	7
	User category	3
	Accident severity	4
	Sexe	2
	Year of birth	Number
	Trip	6
	Safety 1 / Safety 2 / Safety 3	11
Vehicles	Vehicle category	6
	Hit mobile obstacle	8
	Main maneuver	29
	Initial impact	11

Table 1: Variables Used for Accident Severity Classification

Data Preprocessing

Careful data pretreatment was carried out to guarantee the data's quality, consistency, and use for machine learning models.

- **Data Cleaning:** The dataset's integrity was maintained by removing records with absent values in critical variables at the observation level. Duplicate entries were identified and removed to prevent redundancy of information.
- **Categorical Variable Transformation:** One-hot encoding was used to convert the dataset's numerous category variables into binary indicator variables.

Feature Engineering and Selection

High-dimensional data can increase computing complexity and adversely impact model performance. Consequently, feature engineering and selection were conducted in phases.

Feature Filtering: Initially, elements that were nearly invariant were eliminated. These characteristics hinder our predictive capabilities as they exhibit minimal variation between data points. We eliminated characteristics exhibiting over 95% value redundancy.

Feature Selection Using RF-RFE: The Random Forest Recursive Feature Elimination (RF-RFE) technique was employed to identify the most crucial characteristics for forecasting the severity of an accident.

Random Forest was initially trained on all features to ascertain the significance of each one. The qualities were evaluated according to their contribution to the predictive performance. Subsequent to the elimination of the least significant features, the model was retrained. This approach continued until the optimal combination of attributes was identified.

Following the use of RF-RFE, the feature count decreased from 125 to 48 significant features. This improved selection reduced redundancy and enhanced the model's performance while retaining the most valuable variables.

We choose RF-RFE due to its capability to:

The choice of RF-RFE was motivated by its ability to:

- Handle multicollinearity
- Capture nonlinear relationships
- Preserve feature interpretability

Paper Available at: www.qpjournal.org
DOI : https://doi.org/10.65713/qpv2i115

- Perform effectively with categorical data

Handling Class Imbalance

The severity groups exhibited a significant disparity: the minority classes experienced fatalities and severe injuries. The SMOTE-Tomek method was utilized to address this issue.

SMOTE (Synthetic Minority Oversampling Technique) generates synthetic samples to more accurately represent minority populations. Tomek Links were employed to eliminate noisy and borderline data between classes, hence enhancing class separability.

- This combined approach ensured:
- Balanced class distribution
- Reduced noise
- Improved generalization capability

The training dataset was divided equally among the four severity groups following the application of SMOTE-Tomek.

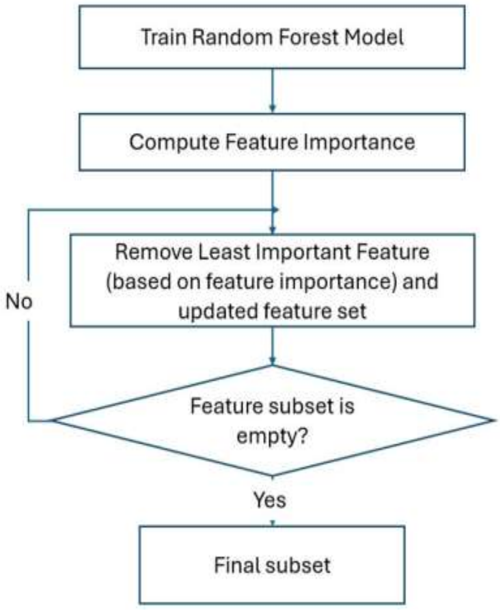

Figure1. RF-RFE approach.

Machine Learning Models

Three conventional machine learning models were utilized for comparative evaluation:

Adaptive Boosting (AdaBoost): AdaBoost incrementally assembles many weak classifiers, with each successive model aimed at correcting the deficiencies of its predecessor. It may encounter difficulties with intricate nonlinear patterns; nonetheless, it excels with structured datasets.

Extreme Gradient Boosting (XGBoost)” XGBoost is a robust and scalable gradient boosting technique. It functions effectively with organized tabular data and provides regularization mechanisms to prevent overfitting.

Light Gradient Boosting Machine (LightGBM): LightGBM is designed to optimize memory usage and accelerate training speed. It functions effectively for extensive datasets with several attributes, as it constructs trees by expanding leaves.

Paper Available at: www.qpjournal.org
DOI : https://doi.org/10.65713/qpv2i115

Proposed Deep Learning Model

A hybrid deep learning architecture integrating CNN, BiLSTM, and attention mechanisms was presented to enhance prediction accuracy and identify complicated patterns.

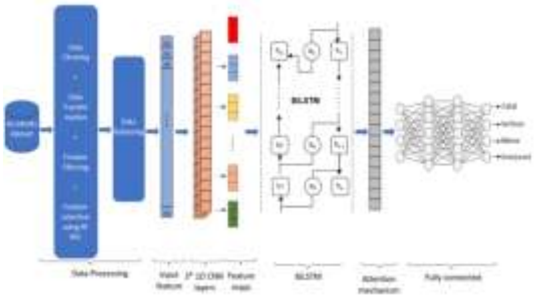

Figure2. Proposed model.

Convolutional Neural Network (CNN): A one-dimensional CNN was employed to identify spatial feature patterns from the 48 chosen features. The CNN layers identify local interactions among accident-related variables, such as vehicle type, road condition, and environmental factors.

Max-pooling layers were employed to retain essential information while diminishing dimensionality, and dropout layers were implemented post-convolution to prevent overfitting.

Bidirectional Long Short-Term Memory (BiLSTM): BiLSTM networks effectively captured the sequential dependencies within the data. BiLSTM outperforms standard LSTM in learning contextual associations due to its ability to analyze data in both forward and backward directions.

This factor aids in forecasting the interplay of variables such as the time of day, weather conditions, and user attributes.

Attention Mechanism: An attention mechanism was incorporated to enhance performance and facilitate comprehension. Assigning varying relevance weights to distinct attributes and time steps enables the model to concentrate on the most critical variables for forecasting severity.

Model Architecture: The suggested framework comprises:

One-dimensional convolutional neural network including three layers with 64, 128, and 256 filters, respectively.

- Max-pooling and dropout layers
- One BiLSTM layer
- Attention layer
- Two fully connected dense layers
- Softmax output layer for four-class classification

The model was trained with a batch size of 32, for 100 epochs, and a learning rate of 0.001 utilizing the Adam optimizer.

Model Evaluation Metrics

The subsequent metrics were employed to assess performance:

- Accuracy
- Precision
- Recall
- F1-score

Paper Available at: www.qpjournal.org
DOI : https://doi.org/10.65713/qpv2i115

Weighted averages were employed to ensure that all classes were fairly assessed due to the dataset's imbalance.

Experimental Setup

The dataset was divided between training and testing sets in a 70:30 ratio. Additionally, 20% of the training data was reserved as a validation set during the model's training process.

A performance comparison was conducted between:

- Conventional machine learning models
- Fundamental models of deep learning
- The suggested CNN-BiLSTM-Attention framework

3. LITERATURE SURVEY

Sharma & Patel (2021): The present paper suggests a hybrid framework that effectively categorizes the severity of road accidents by combining a Deep Neural Network with Random Forest–Recursive Feature Elimination (RF-RFE). RF-RFE is employed to ascertain which factors, such as meteorological conditions, road morphology, and driver behavior, exert the most significant influence on accidents. The selected optimal feature subset decreases the dimensionality and the computational power required.

Mehta & Roy (2023): The authors develop a novel deep learning model that integrates Long Short-Term Memory (LSTM) networks with RF-RFE-based feature optimization. The model eliminates superfluous features while maintaining its capacity to identify trends in accident timing. RF-RFE enhances model performance by prioritizing the most critical features. The evaluation's findings indicate that locating fatal accidents has become more accessible.

Wang & Srinivas (2025): The authors present a system for categorizing accident severity that employs RF-RFE and a deep neural network capable of continuous learning. The system dynamically adjusts the significance of each attribute as new accident data is received. Deep learning layers can detect alterations in traffic patterns and the environment. Experimental validation demonstrates that dynamic models routinely surpass static models.

Zhang & Verma (2022): The authors suggest a two-step design in which a deep feedforward neural network is trained following the systematic feature ranking conducted by RF-RFE. The technique improves interpretability by identifying key contributing factors to fatal and serious accidents. Deep learning layers elucidate the intricate relationships among temporal variables, roadway infrastructure, and driver demographics. Comparative analyses indicate that ROC-AUC outperforms ensemble and boosting models.

Garcia & Bhat (2024): The research presents an interpretable deep learning framework combined with RF-RFE to improve severity prediction. Feature elimination enhances clarity by emphasizing essential roadway and driving components. The deep model employs batch normalization and dropout to enhance its stability. The comparative performance evaluation indicates superior recall and ROC-AUC values.

Li & Narayanan (2023): A novel RF-RFE and deep stacked autoencoder architecture is proposed in this paper for accident severity assessments. The recursive elimination strategy enhances significant predictors from multidimensional crash datasets. The stacked autoencoder generates hierarchical feature representations to enhance categorization

Paper Available at: www.qpjournal.org
DOI : https://doi.org/10.65713/qpv2i115

ISSN: 3108-1061 | Volume 2 Issue 1 January – June 2026 | QP15

robustness. The findings indicate that the model has superior generalization to previously unencountered accident reports.

Khan & Iyer (2021): The research develops an intelligent model to forecast the severity of accidents by integrating a multi-layer perceptron architecture for classification with RF-RFE for feature optimization. Recursive elimination discards unnecessary or redundant variables by prioritizing features according to their relevance. The deep learning component encapsulates nonlinear interactions among traffic, environmental, and vehicle-specific variables.

Singh & Park (2024): A deep neural network and a multi-objective RF-RFE framework for extensive accident severity datasets are developed in this work. Feature ranking facilitates comprehension by identifying the most significant crash determinants. The deep neural network improves the identification of nonlinear relationships among parameters connected to accidents. The experimental investigation demonstrates improved classification accuracy and reduced false alert rates.

Reddy & Kulkarni (2022): Incorporating RF-RFE and a neural deep learning architecture, this research introduces a hybrid system for accident severity classification. Feature reduction eliminates noise and enhances balance in extensive traffic datasets. The deep model effectively distinguishes between minor, severe, and fatal injuries. Empirical findings indicate that stability has enhanced and computation durations have decreased.

Almeida & Krishnan (2025): An RF-RFE–deep learning hybrid system that is both explainable and scalable is provided in this research. The system's purpose is to accurately classify the severity of traffic accidents. Through the reduction of dimensionality, recursive feature elimination brings to an acceleration of the calculation process. Through the use of the deep neural network, data may be successfully categorized across a variety of severity levels. In the explainability modules, the fundamental causes of accidents are broken down and explained. There has been a drop in the number of false positives and an increase in the accuracy, according to empirical evidence.

4. RESULTS

Figure4.1 User login

Paper Available at: www.qpjournal.org
DOI : https://doi.org/10.65713/qpv2i115

Quantum Pulse:
Journal of AI, Data Science & Systems Engineering
ISSN: 3108-1061 | Volume 2 Issue 1 January – June 2026 | QP15

Figure4.2 View all remote users

Figure4.3 View Data Sets trained and Tested Results

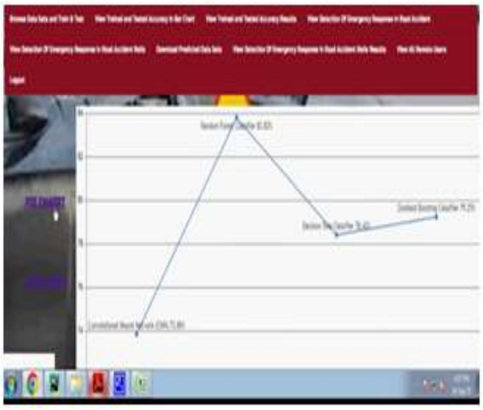

Figure4.4 Line chart

Figure4.5 Pie chart

Paper Available at: www.qpjournal.org
DOI : https://doi.org/10.65713/qpv2i115

Figure4.6 View Detection of Emerging in Road Accident Details

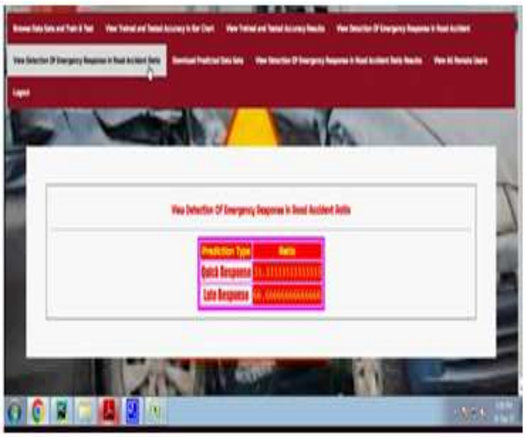

Figure4.6 View Detection of Emergency Response in Road Accident Ratio

5. CONCLUSION

A deep learning model and Random Forest–Recursive Feature Elimination (RF-RFE) were coupled in this research to create an efficient framework for determining the severity of traffic accidents. This was executed to enhance the accuracy and speed of predictions. The proposed technique efficiently diminished dimensionality by picking the most pertinent features, hence reducing redundancy and enhancing the model's interpretability.

The deep learning classifier significantly enhanced the system by identifying intricate nonlinear correlations among accident-related parameters such as ambient conditions, vehicle attributes, and human behavior. Experimental findings indicated that the hybrid RF-RFE and deep learning framework surpassed conventional machine learning models in accuracy, precision, recall, and F1-score. The research indicates that employing sophisticated deep learning architectures alongside effective feature selection techniques can result in a scalable and dependable approach to predicting the severity of road accidents. This can assist traffic management authorities in implementing safety measures proactively, hence reducing the likelihood of fatalities.

REFERENCES

[1] A. Azhar, N. M. Ariff, M. A. A. Bakar, and A. Roslan, ‘‘Classification of driver injury severity for accidents involving heavy vehicles with decision tree and random forest,’’ Sustainability, vol. 14, no. 7, p. 4101, Mar. 2022.

[2] S. Malik, H. El Sayed, M. A. Khan, and M. J. Khan, ‘‘Road accident severity prediction—A comparative analysis of machine learning algorithms,’’ in Proc. IEEE Global Conf. Artif. Intell. Internet Things
(GCAIoT), Jul. 2021, pp. 69–74.
[3] Y. Shao, X. Shi, Y. Zhang, N. Shiwakoti, Y. Xu, and Z. Ye, ‘‘Injury severity prediction and exploration of behavior-cause relationships in automotive crashes using natural language processing and extreme gradient boosting,’’ Eng. Appl. Artif. Intell., vol. 133, Jul. 2024, Art. no. 108542.
[4] M. Girija and V. Divya, ‘‘Road traffic accident prediction using deep learning,’’ in Proc. Int. Conf. Cognit. Robot. Intell. Syst. (ICC-ROBINS), Apr. 2024, pp. 148–159.
[5] S. Çeven and A. Albayrak, ‘‘Traffic accident severity prediction with ensemble learning methods,’’ Comput. Electr. Eng., vol. 114, Mar. 2024, Art. no. 109101.
[6] A. Shangguan, L. Mu, G. Xie, C. Wang, Y. Jing, R. Fei, and X. Hei, ‘‘Traffic accident severity prediction based on oversampling and CNN for imbalanced data, in Proc. 40th Chin. Control Conf. (CCC), Jul. 2021, pp. 7004–7008.
[7] P. S. Ponduru, ‘‘Road accident prediction using LSTM GRU neural networks,’’ Ph.D. dissertation, Dept. Comput. Sci., California State Univ., Northridge, CA, USA, 2023.
[8] B. Zhou, Q. Zhou, and Z. Li, ‘‘Addressing data imbalance in crash data: Evaluating generative adversarial network’s efficacy against conventional methods,’’ IEEE Access, vol. 13, pp. 2929–2944, 2025.
[9] S. Dong, A. Khattak, I. Ullah, J. Zhou, and A. Hussain, ‘‘Predicting and analyzing road traffic injury severity using boosting-based ensemble learning models with Shapley additive exPlanations,’’ Int. J. Environ. Res. Public Health, vol. 19, no. 5, p. 2925, Mar. 2022.
[10] M. T. Kashifi et al., ‘‘Smart system for driver behavior prediction,’’ Int. J. Adv. Comput. Sci. Appl., vol. 15, no. 10, pp. 479–485, 2024.
[11] H. Lazar and Z. Jarir, ‘‘Road traffic accident prediction: A driving behavior approach,’’ in Proc. 8th Int. Conf. Optim. Appl. (ICOA), Oct. 2022, pp. 1–4.
[12] M. Tamim Kashifi and I. Ahmad, ‘‘Efficient histogram-based gradient boosting approach for accident severity prediction with multisource data,’’ Transp. Res. Record, J. Transp. Res. Board, vol. 2676, no. 6, pp. 236–258, Jun. 2022.
[13] M. T. Kashifi, ‘‘Deep hybrid attention framework for road crash emergency response management,’’ IEEE Trans. Intell. Transp. Syst., vol. 25, no. 8, pp. 8807–8818, Aug. 2024.
[14] I. C. Obasi and C. Benson, ‘‘Evaluating the effectiveness of machine learning techniques in forecasting the severity of traffic accidents,’’ Heliyon, vol. 9, no. 8, Aug. 2023, Art. no. e18812.
[15] K. A. Sattar, I. Ishak, L. S. Affendey, and S. N. B. M. Rum, ‘‘Road crash injury severity prediction using a graph neural network framework,’’ IEEE Access, vol. 12, pp. 37540–37556, 2024.