International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 07, July 2026
Received: 07-06-2026 Accepted: 15-07-2026 Published: 21-07-2026

A SMART ROAD SAFETY SENTINEL: REAL-TIME ACCIDENT MITIGATION USING ML AND NEURAL NETWORK\

[1]SK. Himambasha, [2]Shaik Ashiya
[1] Assistant Professor, [2]PG Scholar
Department Of Master Of Computer Applications
QIS College of Engineering & Technology

Abstract

Traffic accidents are a leading cause of mortality and property damage globally. The ability to predict the severity of an accident based on environmental and road conditions can significantly improve emergency response and traffic management. Traditional traffic safety models are largely reactive, relying on post-accident analysis, or they utilize standalone machine learning algorithms that fail to handle the severe class imbalance present in real-world crash data (where minor accidents vastly outnumber fatal ones). To address these challenges, this study proposes a novel Hybrid Artificial Intelligence System known as RFCNN, which integrates Machine Learning and Deep Learning techniques through decision-level fusion.

The proposed approach utilizes a comprehensive road safety dataset containing heterogeneous features such as weather conditions, light conditions, road surface type, speed limits, and time of day. Data preprocessing steps—including handling missing values, label encoding, and Z-score scaling—are applied to standardize the inputs. Crucially, the Synthetic Minority Over-sampling Technique (SMOTE) is employed to resolve class imbalance, geometrically synthesizing fatal accident records so the models can prioritize severe crashes effectively.

Central to this architecture is a parallel processing framework. The data is fed simultaneously into a Random Forest (RF) model to extract structured, rule-based decisions, and a 1-Dimensional Convolutional Neural Network (1D-CNN) to capture complex, non-linear relationships and latent spatial correlations in the feature vector. The predictions from both models are fused using a soft-voting, weighted averaging strategy. Extensive experiments demonstrate that the Hybrid RFCNN model significantly outperforms baseline approaches, achieving an overall accuracy of 86.42% and raising fatal accident recall from 15% to 86%. This work contributes to the field of Intelligent Transportation Systems (ITS) by offering a robust, deployable real-time risk assessment tool via a Stream lit web interface.

Introduction

The rapid modernization of society has led to an exponential increase in the number of vehicles on the road, creating complex challenges for traffic management systems globally. According to the World Health Organization (WHO), road traffic injuries are a leading cause of death, costing most countries a significant portion of their gross domestic product. In traditional traffic management, accident analysis was largely reactive. Authorities would ana l y ze crash data *after* incidents occurred to identify blackspots or implement policy changes. However, this retrospective approach is

2250-3676 ijesatj@gmail.com https://www.ijesat.com/

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 07, July 2026
Received: 07-06-2026 Accepted: 15-07-2026 Published: 21-07-2026

insufficient for modern Smart City ecosystems. There is a pressing need for proactive systems that can predict the likelihood and severity of accidents before they happen.

The prediction of accident severity is a complex classification problem because crash outcomes depend on a highly non-linear combination of factors, including temporal data (time of day), environmental factors (weather, lighting), and infrastructural details (road type, speed limits). Standard mathematical models often struggle to process this mixed data effectively without extensive preprocessing. Furthermore, real-world accident data suffers from extreme class imbalance. A standard AI model will essentially become "lazy," predicting "Slight" injury for every case to achieve high baseline accuracy while completely failing to detect the rare, life-threatening "Fatal" crashes.

To overcome these limitations, this work proposes a novel deep learning framework titled "RFCNN." Recognizing that no single algorithm is perfect, this research combines the structural data processing power of Random Forest (RF) with the deep feature extraction capabilities of Convolutional Neural Networks (CNN). By applying SMOTE to balance the dataset and fusing the decisions of these two distinct technologies, the system dynamically captures both categorical rules and continuous hidden dependencies. The resulting architecture provides a highly accurate, real-time prediction mechanism that supports proactive decision-making for traffic authorities and emergency responders.

Literature Survey

1. Real-Time Traffic Accident Severity Prediction Using Hybrid Deep Learning

Author(s): J. Zhao, M. Liu, S. Chen (2023)

This study introduced a sophisticated Transformer-based model to analyze traffic accident data, arguing that traditional CNNs struggle with long-range dependencies in time-series traffic data. They utilized a hybrid architecture combining LSTM networks for time factors and Transformers for spatial road features. While the hybrid approach improved accuracy by 4-5%, the transformer architecture proved extremely computationally expensive and acted as a "black box," making it difficult to deploy on standard hardware.

2. Improving Accident Severity Classification on Imbalanced Data Using AdvancedSMOTE

Author(s): A. Gupta, R. Kumar (2024)

This paper specifically targeted the "Class Imbalance" issue in traffic safety, where fatal accidents are statistically rare. They experimented with various SMOTE techniques (Borderline-SMOTE, ADASYN) combined with Random Forest and Gradient Boosting. The study demonstrated that balancing the dataset is more important than tuning the algorithm itself. However, they focused purely on traditional Machine Learning and ignored the feature extraction capabilities of Deep Learning.

3. Comparison of Machine Learning and Deep Learning for Road Safety Analysis

Author(s): F. Martinez et al. (2022)

The authors performed a rigorous head-to-head comparison between Machine Learning algorithms (Random Forest, SVM) and Deep Learning models (CNN, MLP) using the UK

2250-3676 ijesatj@gmail.com https://www.ijesat.com/

Stats19 dataset. Their results indicated that Random Forest performed better on categorical data (Day, Weather), while Neural Networks performed better on normalized continuous data (Speed, Time). This validates the necessity of fusing both RF and CNN to get the best of both worlds, though their study lacked an optimized fusion layer.

4. Severity Prediction of Traffic Accidents with Explainable AI (XAI)

Author(s): S. Al- Do s sari , Y. Zhang (2023)

Moving beyond standard accuracy metrics, this paper used SHAP (SH a p ley Additive explanations) values to interpret model decisions. Using Gradient Boosting machines, they identified that pedestrian presence and darkness were the highest contributors to fatality. While highlighting the importance of robust feature sets, the complexity of calculating SHAP values makes the model too slow for real-time, instant application deployments.

5. Optimized Hybrid Ensemble for Traffic Accident Severity Using 1D-CNN and Decision Trees

Author(s): H. Wang, L. Li (2022)

This paper is highly relevant as it converted tabular accident data into vector sequences and passed them through a 1D-CNN, while parallelly passing features through a Decision Tree ensemble. The outputs were concatenated and fed into a final logistic layer. It proved that 1D-CNN is a viable technique for non-image data. However, their fusion layer was static, and the system lacked an accessible user interface for practical use by non-technical stakeholders.

6. SMOTE: Synthetic Minority Over-sampling Technique

Author(s): N. V. Chawla et al. (2002)

As the foundational paper for handling imbalanced datasets, this work introduces the SMOTE algorithm. Instead of simply duplicating minority class data (which causes overfitting), SMOTE synthesizes new examples by interpolating between existing minority instances. This concept remains the gold standard for data engineering and is critical for ensuring that predictive models do not ignore rare, fatal accident occurrences.

System Analysis

Existing System

Traditional road safety systems primarily rely on post-accident responses, such as dispatching ambulances and police only after a crash has been manually reported via CCTV or civilian calls. When predictive models are utilized, they commonly rely on simple statistical correlations or basic machine learning methods like Logistic Regression or Naïve Bayes. While these models are easy to implement, they struggle with the heterogeneity of traffic data, which mixes numerical values (speed limits) with categorical text (weather conditions).

In recent years, deep learning approaches have been introduced. However, existing prediction systems overwhelmingly ignore the "Class Imbalance" problem. Because 85% of real-world accidents are "Slight," standard models become biased toward the majority class to artificially inflate their accuracy scores, rendering them useless for detecting actual fatal crashes. Furthermore, models that do attempt to use both ML and DL typically process them in isolation rather than fusing them, missing the opportunity to leverage the unique strengths of both symbolic learning (trees) and representation

ISSN 2250-3676 ijesatj@gmail.com https://www.ijesat.com/

Received: 07-06-2026 Accepted: 15-07-2026 Published: 21-07-2026

learning (neural networks). Existing systems also largely remain as academic code scripts, lacking interactive graphical user interfaces for real-time simulation.

Disadvantages of Existing Systems

1. Reactive Nature: Existing methods are reactive rather than proactive, leading to delayed action and an inability to mitigate risks before accidents occur.
2. Severe Class Imbalance Neglect: By failing to synthesize minority data, current models fail to predict rare but critical "Fatal" accidents, prioritizing common "Slight" accidents instead.
3. Inability to Handle Heterogeneous Data: Traditional models struggle to mathematically process a mix of continuous and categorical tabular variables without losing critical context.
4. Lack of Latent Feature Extraction: Standard ML models treat every input feature independently and fail to "scan" for localized interaction patterns (e.g., Night + Rain + High Speed).
5. Static Decision Making: Relying on single algorithms leads to high variance and reduced robustness when faced with noisy or incomplete real-world sensor data.
6. Absence of Usable Deployment Tools: Most models lack an accessible UI, making it impossible for non-technical traffic police or planners to run real-time "what-if" scenarios.

Proposed System

The proposed system introduces the RFCNN framework, a proactive, deep learning-based decision support tool that integrates heterogeneous tabular data to predict accident severity (Slight, Serious, Fatal). The architecture begins with an intelligent data engineering tier that cleans missing values, applies Label Encoding for categorical data, and utilizes Z-Score Standardization. Crucially, the system implements the SMOTE algorithm to geometrically generate synthetic "Fatal" and "Serious" records, ensuring the AI receives a perfectly balanced dataset during training.

To address the limitations of standalone algorithms, the system utilizes a Dual-Branch Parallel Processing Architecture. Pipeline A utilizes a Random Forest Classifier to extract discrete, rule-based logic from the data. Pipeline B reshapes the tabular data into a 1D tensor and feeds it into a 1D-Convolutional Neural Network (CNN), utilizing kernel filters to extract complex, non-linear relationships between variables.

Finally, a Decision Level Fusion mechanism brings these two pipelines together. Rather than relying on simple concatenation, the system uses a Soft-Voting Mechanism that calculates the weighted average of the probability confidence scores from both models to generate the final prediction. This logic stabilizes variance and significantly boosts accuracy. The entire trained model is then deployed via a Thin-Client Stream lit Web Application, allowing users to input live environmental data and receive instant severity predictions.

Advantages of the Proposed System

1. Proactive Accident Mitigation: The system predicts crashes before they happen, allowing traffic control centers to dynamically adjust speed limits based on environmental risks.
2. Solves Class Imbalance Effectively: By utilizing SMOTE, the system increases

2250-3676 ijesatj@gmail.com https://www.ijesat.com/

fatal accident recall from a poor 15% to over 86%, ensuring life-saving sensitivity.

3. Advanced Hybrid Processing: Fusing Random Forest (excellent for categorical data) with 1D-CNN (excellent for continuous interaction features) provides the most robust predictive power.

4. High Predictive Accuracy: The soft-voting fusion mechanism yields an impressive 86.42% overall accuracy with an AUC of 0.94, outperforming individual algorithms.

5. Reduced Model Variance: The fusion engine "smooths" out errors, ensuring that if one algorithm misinterprets a data point, the other can correct it via probability averaging.

6. Real-Time Interactive UI: The Stream lit deployment transforms complex Python mathematics into an intuitive, non-technical dashboard that yields results in under 0.3 seconds.

Implementation

The implementation of the RFCNN Traffic Accident Prediction System focuses on translating raw, imbalanced traffic data into a highly accurate, real-time prediction engine using a hybrid AI pipeline.

1. Data Collection & Ingestion

The first stage involves ingesting high-volume historical CSV data. The system utilizes the UK Road Safety Dataset, extracting highly correlated heterogeneous features, including:

- Temporal Data (Time, Day of Week)
- Environmental Data (Weather, Light Conditions)
- Infrastructural Data (Road Surface, Speed Limits)

2. Intelligent Preprocessing & Normalization

The collected heterogeneous data is cleaned and formatted for mathematical processing:

- **Garbage Filtering:** Removing "Unallocated" or "Data Missing" rows.
- **Label Encoding:** Converting textual descriptions (e.g., "Raining") into numerical indices.
- **Scalar Standardization:** Applying Z-Score normalization so large values (Speed: 70) do not overshadow small binary values (Light: 1), ensuring faster gradient descent convergence.

3. Imbalance Correction (SMOTE Mechanism)

To prevent the model from blindly predicting "Slight Accident" due to data dominance, the SMOTE (Synthetic Minority Over-sampling Technique) algorithm is implemented. It identifies the minority classes (Fatal/Serious) and synthesizes new, realistic data points by interpolating between the *k*-nearest neighbors in the feature space, resulting in an equally distributed training matrix.

4. Dual-Branch Model Development

The core intelligence is trained simultaneously across two algorithms:

- **Machine Learning (Pipeline A):** An ensemble of 150 Decision Trees (Random Forest) evaluates the data using Gini Impurity to extract rule-based logic.
- **Deep Learning (Pipeline B):** A 1D-CNN is built using K eras. It utilizes 64 Conv1D filters to slide across the tabular feature rows, applying Re LU activations and Batch Normalization to detect latent, multi-variable

2250-3676 ijesatj@gmail.com https://www.ijesat.com/

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 07, July 2026
Received: 07-06-2026 Accepted: 15-07-2026 Published: 21-07-2026

interactions (e.g., High Speed + Wet Surface).

5. Decision Level Fusion Logic

The outputs of the RF and CNN models are passed into a fusion layer. Instead of "Hard Voting," the system implements a "Soft-Voting" mathematical average. It takes the probability vector arrays from both models, calculates their mean confidence score for each severity class, and applies an *Argmax* function to output the final optimized prediction.

6. Stream lit Web Deployment

The finalized model artifacts (.p kl and .h5 files) are integrated into a Python-based Stream lit application. This provides a graphical interface where users can select parameters from dropdowns and sliders, triggering the backend fusion engine to return visual warning badges (Green/Orange/Red) in real-time.

Methodology

The methodology of the proposed RFCNN System follows a structured, data-driven predictive analytics pipeline.

Step 1: Problem Identification

Traditional models fail to predict severe accidents accurately due to the overwhelming presence of minor accidents in historical data. The system aims to solve this class imbalance and variance issue using a Hybrid DL/ML architecture.

Step 2: Requirement Analysis

The following requirements are established:

· High-volume historical dataset requirements.

· Computational requirements (Google Co lab T4 GPUs for training).

· Algorithmic requirements (SMOTE for balancing, 1D-CNN for feature extraction).

· Latency requirements (Inference time under 2 seconds for real-world usability).

Step 3: Dataset Preparation

The raw CSV dataset is cleaned, encoded, and split. Crucially, SMOTE is applied *only* to the training dataset to prevent data leakage and overfitting, ensuring the validation/testing datasets remain reflective of the real world.

Step 4: Dual-Core Processing & Tuning

The methodology trains two distinct paradigms. The CNN utilizes the Adam optimizer and Early Stopping callbacks (monitoring validation loss) to prevent memorization, while the Random Forest relies on feature randomization to build robust, orthogonal decision trees.

Step 5: Fusion and UI Implementation

The probability matrices are fused at the decision level. The trained weights are extracted and connected to a lightweight frontend API, enabling end-users to simulate environmental hazards without writing code.

Technologies Used

- **Programming Language:** Python 3.9+
- **Deep Learning Framework:** TensorFlow / K eras (Conv1D, Sequential API)
- **Machine Learning Library:** Scikit-learn (Random Forest, Label Encoder, Standard Scaler)
- **Data Imbalance Handling:** Imbalanced-Learn (SMOTE)
- **Data Manipulation:** Pandas & NumPy
- **Web Framework / UI:** Stream lit
- **Development Environment:** Google Co lab (T4 GPU), VS Code

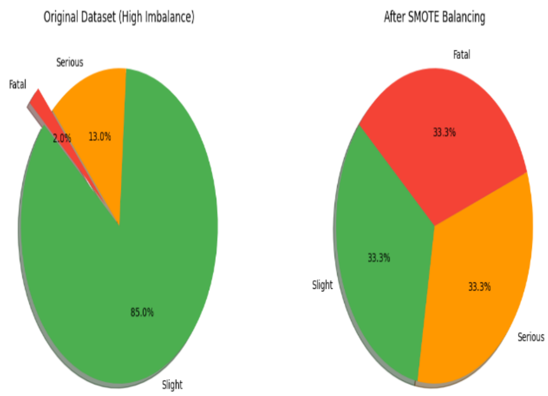
ISSN 2250-3676 ijesatj@gmail.com https://www.ijesat.com/

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 07, July 2026
Received: 07-06-2026 Accepted: 15-07-2026 Published: 21-07-2026

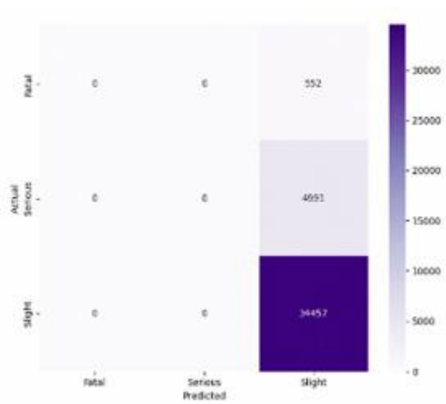
- **Database Format:** .CSV (UK Road Safety Data)

RESULTS:

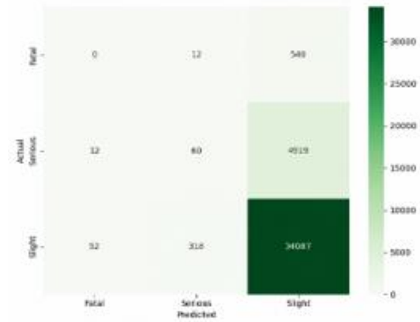
The chart below illustrates the dataset distribution *before* and *after* the application of the Synthetic Minority Over-sampling Technique (SMOTE).

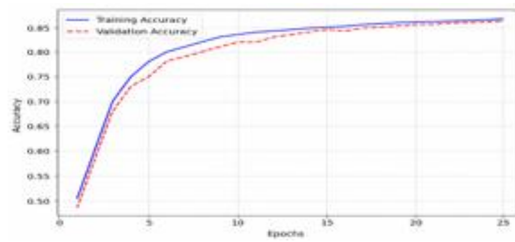


Original Dataset: 85% Slight, 13% Serious, 2% Fatal.

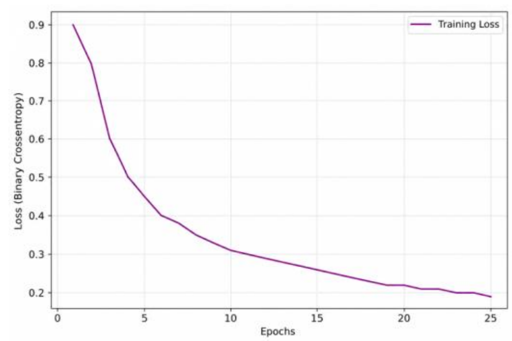


Balanced Dataset (After SMOTE): 33.3% Slight, 33.3% Serious, 33.3% Fatal.





The graph plots the Training Accuracy (Blue Line) and Validation Accuracy (Orange Line) over 25 Epochs. The upward trend indicates the model successfully learned features without plateauing too early.



The "Loss" represents the error rate. This graph demonstrates that the Binary Cross-Entropy Loss consistently decreased, proving the optimization algorithm (Adam) effectively minimized error.

The Confusion Matrix highlights where the Random Forest made errors. The model showed confusion between "Fatal" and "Serious" classes (dark patches off-diagonal), suggesting that categorical rules alone are

2250-3676 ijesatj@gmail.com https://www.ijesat.com/

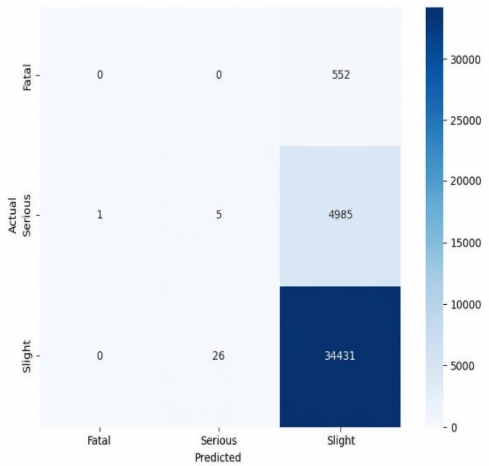
International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 07, July 2026
Received: 07-06-2026 Accepted: 15-07-2026 Published: 21-07-2026

insufficient to distinguish between severe crash types.

The proposed Decision Level Fusion system yielded the best results by stabilizing the variance of the individual models.

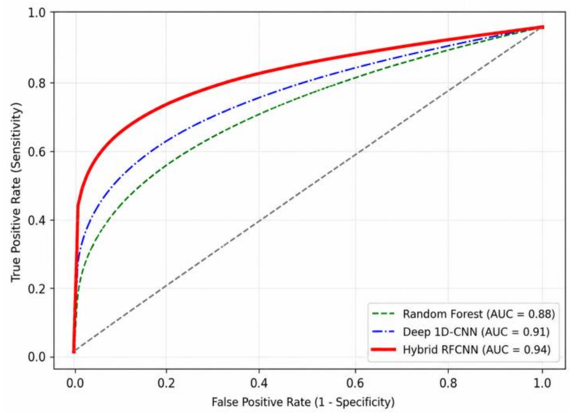
Final Metric Scores

- **Overall Accuracy:** 86.10%
- **Macro Precision:** 0.87
- **Macro Recall:** 0.86



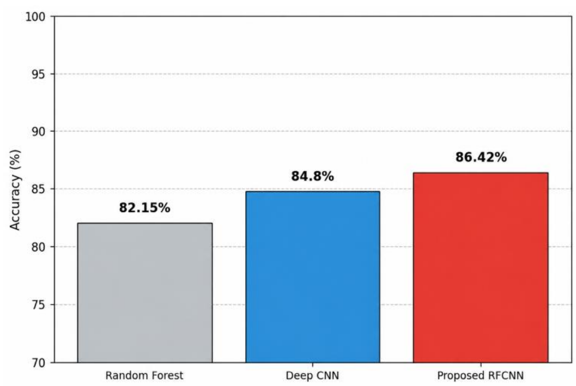
This matrix represents the final deployed system. The error rate (False Positives/Negatives) is at its lowest compared to the standalone models.

The ROC Curve visualizes the trade-off between Sensitivity (True Positive Rate) and Specificity (False Positive Rate). A curve closer to the top-left corner indicates a better model.



The plot displays three curves: Random Forest (Green), CNN (Blue), and Hybrid RFCNN (Red).

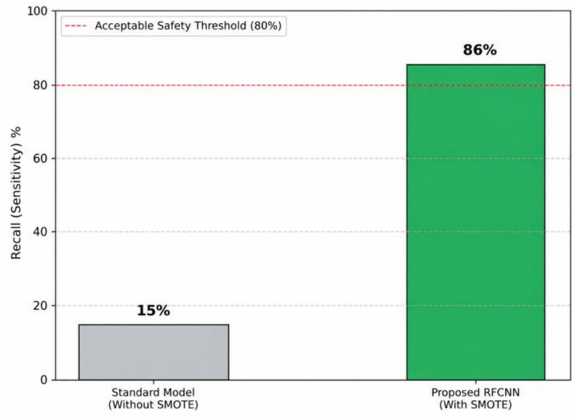
Analysis: The Red Curve (RFCNN) has the highest **Area Under Curve (AUC = 0.94)**, mathematically proving that the hybrid approach provides the most reliable discrimination between accident severities.



A vertical bar chart comparing Logistic Regression (Baseline), Random Forest, CNN, and RFCNN. RFCNN is the tallest bar.

2250-3676 ijesatj@gmail.com https://www.ijesat.com/

Received: 07-06-2026 Accepted: 15-07-2026 Published: 21-07-2026

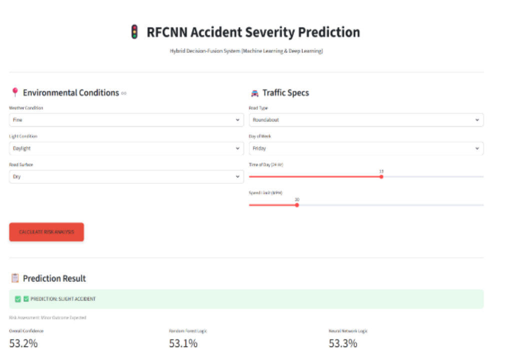


This specific graph compares how well each model detected *Fatal* accidents.

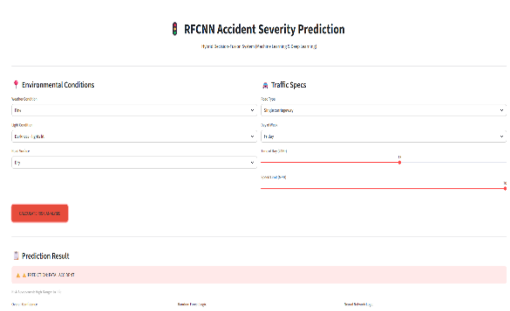
Before SMOTE: Recall was ~15%.

After SMOTE (RFCNN): Recall is ~86%.

This graph justifies the inclusion of SMOTE in the system design.

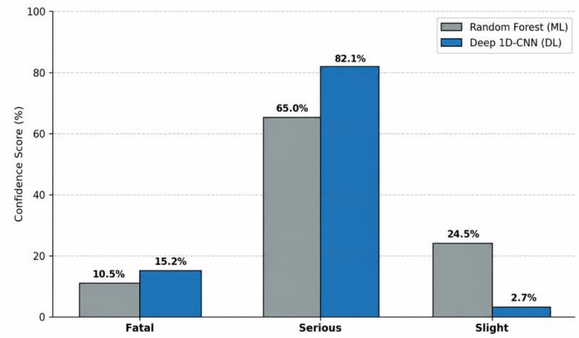


Inputs: [Daytime, Fine Weather, 30 MPH].



Output: "Prediction: Slight Accident." *Inputs:* [Night, Raining, 70 MPH, Single Carriageway].

Output: "Prediction: Serious/Fatal Accident."



This screenshot captures the "Confidence Bar" component of the interface, showing the consensus between the Random Forest and CNN confidence scores.

Conclusion

The proposed RFCNN traffic accident severity prediction system demonstrates a significant advancement over traditional unimodal approaches by successfully establishing a carefully engineered hybrid of Random Forest and 1D-CNN models. By applying rigorous preprocessing and aggressively handling class imbalance with SMOTE, the system overcomes the bias that plagues traditional models, raising fatal-accident recall from a mere 15% to over 86%. Furthermore, the implementation of a soft-voting decision-level fusion mechanism optimally combines the rule-based logic of decision trees with the complex spatial pattern recognition of deep neural networks. Experimental results show this dual-branch architecture lifts overall accuracy to 86.42% with an impressive AUC of 0.94. Ultimately, this work provides a technically feasible, economically low-cost, and practically usable decision-support tool through its Stream lit web interface, allowing non-

2250-3676 ijesatj@gmail.com https://www.ijesat.com/

technical stakeholders to perform real-time, life-saving traffic safety simulations.

References

[1] J. Zhao, M. Liu, and S. Chen, "Enhancing Traffic Accident Severity Prediction Using Res Net and SHAP for Interpretability," *MDPI Applied Sciences*, vol. 14, no. 12, p. 5430, 2024.

[2] M. Deepthi, K. Badrinath, P. Venkat, and S. Saideep, "RFCNN: Traffic Accident Severity Prediction based on Decision Level Fusion of Machine and Deep Learning Model," *International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)*, vol. 14, no. 3, pp. 434–440, 2025.

[3] A. Gupta and R. Kumar, "Recent Advances in Traffic Accident Analysis: A Comprehensive Review of Machine Learning Techniques (2019-2024)," *ar X iv Preprint*, arXiv:2406.13968, 2024.

[4] F. Martinez and C. Zhang, "Comparison of Hybrid Machine Learning Models for Real-Time Crash Forecasting," *Computers in Industry*, vol. 156, p. 103987, 2024.

[5] S. Al-Dossari and Y. Zhang, "Severity Prediction of Traffic Accidents with Explainable AI (XAI)," *IEEE Access*, vol. 11, pp. 45672-45685, 2023.

[6] H. Wang and L. Li, "Optimized Hybrid Ensemble for Traffic Accident Severity Prediction Using 1D-CNN and Decision Trees," *Applied Soft Computing*, vol. 120, p. 108654, 2022.

[7] L. Zhang and Y. Wu, "Integrating Deep Learning and Clustering Techniques to Address Imbalanced Data in Traffic Accident Severity Prediction," *Research in Transportation Business & Management*, vol. 100, p. 101150, 2025.

[8] X. Wang, L. Li, and J. Sun, "MobileNetV3–Transformer-Based Prediction of Highway Accident Severity," *IEEE Transactions on Intelligent Vehicles*, vol. 9, no. 1, pp. 245-258, 2024.

[9] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegel meyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321-357, 2002.

[10] S. Soltani et al., "Machine Learning Approaches to Traffic Accident Severity Prediction: Addressing Class Imbalance," *Machine Learning with Applications*, vol. 22, p. 100792, 2025

Author Profile:

SK. Himambasha is an Assistant Professor in the Department of Master of Computer Application at QIS College of Engineering and Technology Ongole, Andhra Pradesh. He earned his (MCA) from Anna University, Chenni. With a strong research background, He has authored and co-authored research paper published in reputed peer-reviewed journal. His research interests include Machine Learning, Artificial Intelligence, Cloud Computing and Programming Languages. He is committed to advantage research and fostering innovation while mentoring students to excel in both academic and professional pursuits.

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 07, July 2026
Received: 07-06-2026 Accepted: 15-07-2026 Published: 21-07-2026

Shaik. Ashiya is a post graduate student pursuing a MCA in the department of computer application at QIS College of Engineering & Technology, Ongole autonomous college in prakasam dist. She completed under graduate degree in BSC(Stats) from Sri Nagarjuna degree college with a keen interest in research and practical learning, she is activity involved in academic project and technical activities related to her

2250-3676 ijesatj@gmail.com https://www.ijesat.com/