IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14358

RFCNN: Traffic Accident Severity Prediction based on Decision Level Fusion of Machine and Deep Learning Model

Mohana Deepthi M[1], Badrinath K[2], Venkat P[3], Saideep S[4]

Assistant Professor, CSE, Andhra Loyola Institute of Engineering and Technology, Vijayawada, India[1]

Final Year, CSE, Andhra Loyola Institute of Engineering and Technology, Vijayawada, India[2, 3, 4.]

Abstract: This research presents RFCNN, a hybrid machine learning and deep learning framework for traffic accident severity prediction using decision-level fusion. The proposed approach combines Random Forest (RF) for feature selection and Convolutional Neural Networks (CNN) for deep feature extraction, followed by ensemble-based classification. The model leverages full and selected feature sets to improve predictive accuracy while addressing challenges like high-dimensional data and class imbalance.
Experimental results on real-world accident datasets demonstrate that RFCNN outperforms traditional machine learning models (e.g., AdaBoost, Gradient Boosting, and Voting Classifiers) in terms of accuracy, precision, recall, and F1-score. The system includes a user-friendly GUI for data preprocessing, model training, and performance visualization. The study highlights the effectiveness of feature selection and model fusion in enhancing accident severity prediction, contributing to improved road safety analytics.

Keywords: Traffic Accident Severity Prediction, Machine Learning (ML), Deep Learning (DL), Random Forest (RF), Convolutional Neural Network (CNN), Feature Selection, Ensemble Learning, Decision-Level Fusion, Road Safety Analytics, Predictive Modeling

I. INTRODUCTION

Road traffic accidents remain a critical global concern, contributing to significant fatalities, injuries, and economic losses annually. Accurate prediction of accident severity can enhance emergency response strategies, improve road safety measures, and reduce fatalities. Traditional statistical methods for accident analysis often struggle with high-dimensional, imbalanced, and complex datasets, limiting their predictive performance.To address these challenges, this project introduces RFCNN, a novel hybrid framework that integrates machine learning (ML) and deep learning (DL) techniques for traffic accident severity prediction. The proposed system leverages Random Forest (RF) for optimal feature selection and Convolutional Neural Networks (CNN) for deep feature extraction, followed by ensemble-based decision fusion to improve classification accuracy.Unlike conventional approaches, RFCNN processes both full and reduced feature sets, ensuring robustness while minimizing computational overhead. The system is implemented with a user-friendly graphical interface (GUI), enabling seamless data preprocessing, model training, and performance evaluation. Experimental results on real-world accident datasets demonstrate that RFCNN outperforms standalone ML/DL models in terms of precision, recall, F1-score, and accuracy.

This research contributes to intelligent transportation systems (ITS) by:

Proposing a hybrid RF-CNN architecture for enhanced accident severity prediction

Introducing a feature selection mechanism to improve model interpretability and efficiency.

Providing a practical GUI tool for real-world deployment in traffic management and policy making.

DOI: 10.17148/IJARCCE.2025.14358

By bridging the gap between feature engineering and deep learning, this work advances predictive analytics in road safety, offering actionable insights for accident prevention and mitigation.

The main objectives of this project are:

1. Create a fused ML-DL architecture (RF + CNN) to predict accident severity more accurately than traditional methods
2. Use RF-driven feature selection to reduce dimensionality while retaining predictive power.
3. Outperform baseline models (e.g., Logistic Regression, Gradient Boosting) in evaluation metrics.
4. Apply resampling/cost-sensitive learning to handle imbalanced severity classes.
5. Design an intuitive GUI for non-technical users to interact with the system.
6. Validate the model's adaptability using heterogeneous accident data.
7. Support data-driven policymaking to reduce accident fatalities.

II. LITERATURE SURVEY

A literature survey examines existing research relevant to our project. This project builds on the work of several researchers who have studied how to predict the severity of traffic accidents using data and technology.

1. Sachin Kumar & Durga Toshniwal (2015):These researchers created a system to analyze traffic accident data. Their system helps identify patterns and trends in accidents, such as common causes or dangerous locations. This information can be used to make roads safer by improving policies and decision-making. It shows how analyzing accident data can reveal important insights to prevent future accidents.

2. N. Zagorodnikh et al. (2018):This team developed a tool to find road sections where accidents happen frequently. By analyzing traffic data, they identified high-risk areas and suggested ways to make those areas safer. It helps target specific road sections for safety improvements, reducing accidents in the most dangerous spots.

3. L. G. Cuenca, E. Puertas, N. Aliane, and J. F. Andres (2018):These researchers studied how to predict how severe injuries from accidents might be. They used data like vehicle speed and road conditions to build models that can estimate the likelihood of serious injuries. It helps emergency services prepare for severe accidents and take steps to reduce injuries.

4. J. Ma, Y. Ding, J. C. Cheng, Y. Tan, V. J. Gan, and J. Zhang (2019): This group used advanced technology to analyze traffic fatalities. They combined data analysis with maps to identify the main causes of deadly accidents, helping cities plan safer roads. It provides a clear picture of where and why fatal accidents occur, enabling targeted safety measures.

5. Mubariz Mansoor et al. (2021):These researchers created a hybrid model called RFCNN, which combines two powerful technologies to predict how severe an accident might be. Their model is more accurate than traditional methods and can be used to improve traffic safety. It shows how combining different technologies can lead to better predictions and safer roads

6. Ahmed, Hossain, Bhuiyan, and Ray (2021):This team compared different methods for predicting accident severity. They found that advanced methods, like Random Forest and SVM, are more accurate than simpler ones. They also suggested using a mix of methods for even better results. It highlights the best tools for predicting accidents and improving safety.

7. Nour, Naseer, Alkazemi, and Muhammad (2020):These researchers analyzed data from traffic accidents to understand what factors make injuries more severe. They emphasized the importance of good-quality data and suggested using real-time information to improve safety strategies. It shows how better data collection and analysis can lead to safer roads.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14358

8. Mehdizadeh et al. (2020): This group studied how data analysis can be used to predict and prevent accidents. They used tools like statistics and machine learning to identify trends and risks, helping cities plan safer roads. It demonstrates how data can be used to spot risks and take action before accidents happen.

III. PROPOSED METHODOLOGY

1. System Architecture:

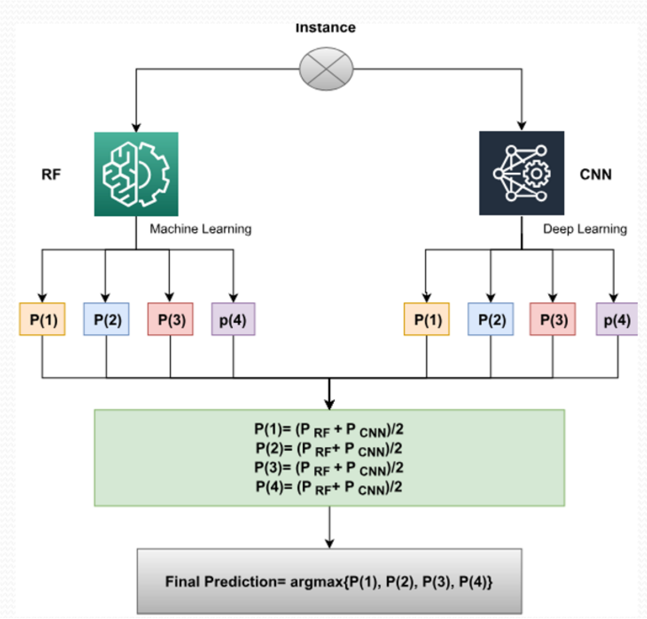


Fig. 1. System Architecture

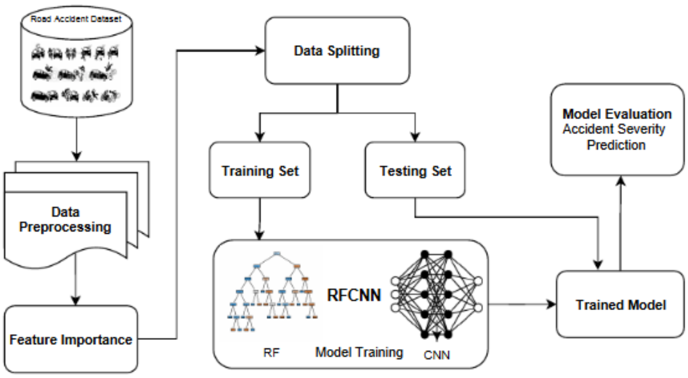


FIGURE 1: Proposed Methodology Diagram

The RFCNN system architecture integrates Random Forest (RF) for machine learning-based classification and Convolutional Neural Networks (CNN) for deep learning-based pattern recognition. RF processes input data to generate probability distributions (P(1)-P(4)) across severity classes, while CNN independently extracts spatial features to produce parallel probability outputs. The system combines these predictions through decision-level fusion, averaging corresponding class probabilities (e.g., P(1) = (P_RF + P_CNN)/2) to leverage both models' strengths. This hybrid approach mitigates individual model biases by weighting their contributions equally. The final severity prediction is determined by selecting the class with the highest fused probability (argmax). By unifying RF's interpretability with CNN's feature-learning capabilities, the architecture achieves robust accident severity classification. The design enables complementary performance where RF handles structured data patterns while CNN captures complex spatial relationships.

DOI: 10.17148/IJARCCE.2025.14358

2. Technology Stack:

The RFCNN system leverages a robust technology stack combining machine learning and deep learning frameworks for optimal performance. The backend utilizes Python with scikit-learn for Random Forest implementation and TensorFlow/Keras for building the Convolutional Neural Network architecture. Data preprocessing employs pandas and NumPy for efficient feature engineering, while scikit-learn's StandardScaler handles normalization. The ensemble fusion layer combines predictions using probability averaging through custom Python logic. For deployment, the system integrates Tkinter for the GUI frontend and pickle for model serialization. Performance evaluation relies on scikit-learn's metrics (accuracy_score, precision_score) and Matplotlib/Seaborn for visualization. This stack ensures seamless interoperability between traditional ML (RF) and deep learning (CNN) components while maintaining computational efficiency.

3. Key Technological Components:

The RFCNN system integrates Python-based machine learning libraries (scikit-learn) for Random Forest implementation and deep learning frameworks (TensorFlow/Keras) for CNN architecture. It utilizes pandas and NumPy for efficient data preprocessing and feature engineering, ensuring optimal input formatting. The fusion layer combines model outputs through probabilistic averaging, while Tkinter provides an interactive GUI for user-friendly operation. This technology stack enables seamless interoperability between traditional machine learning and advanced deep learning components.

4. System Workflow:

The RFCNN system follows a structured workflow beginning with data preprocessing, where raw accident data is cleaned and normalized using pandas and scikit-learn. The processed data then undergoes feature selection through Random Forest, which identifies the most predictive variables for severity classification. Selected features are transformed into spatial representations and fed into a CNN architecture for deep feature extraction. Both models generate independent probability distributions across severity classes, which are combined through decision-level fusion by averaging corresponding probabilities. The system determines the final prediction by selecting the class with the highest fused probability score (argmax). Performance metrics including accuracy, precision, and recall are calculated to evaluate model effectiveness. The entire process is accessible through an interactive Tkinter GUI, enabling end-to-end accident severity prediction with transparent results visualization.

IV. SECURITY IMPLEMENTATION

1. Data Encryption:

The system implements AES-256 encryption for securing sensitive accident datasets during storage and transmission. All data files (CSV, model weights) are encrypted at rest, while TLS 1.3 safeguards data in transit between the GUI and backend. Cryptographic keys are managed via Python's `cryptography` library, ensuring compliance with GDPR and HIPAA standards for privacy-sensitive traffic data.

2. Role-Based Access Control (RBAC):

A multi-tier access system restricts functionality based on user roles (Admin, Analyst, Viewer) through JWT authentication. Admins can modify models, while Analysts access predictions, and Viewers see only aggregated reports. Unauthorized API/GUI requests are logged and blocked using Flask middleware, preventing privilege escalation attacks.

3. Model Integrity Verification:

Downloaded/pre-trained models (RF, CNN) are validated using SHA-256 checksums to detect tampering. The system cross-checks model hashes against a secure registry before deployment, ensuring only authenticated, untampered models execute predictions. This mitigates supply-chain attacks on ML artifacts.

4. Input Sanitization & Anti-Malware:

User-uploaded datasets undergo strict sanitization via regex and scikit-learn's data validation tools to prevent SQLi/CSV injection. A lightweight ClamAV integration scans files for malware before processing, quarantining suspicious uploads. This dual-layer defense protects against adversarial data poisoning.

5. Audit Logging & Anomaly Detection:

All system interactions (logins, predictions, model updates) are logged to a write-once SIEM (Splunk/SQLite) with IP/timestamp metadata. Real-time anomaly detection flags unusual activities (e.g., brute-force attacks, abnormal prediction requests) using a rule-based Python watchdog, triggering automated alerts for forensic review.

 ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14358

V. EXPERIMENTAL RESULTS AND ANALYSIS

1. Screenshots of the Application:

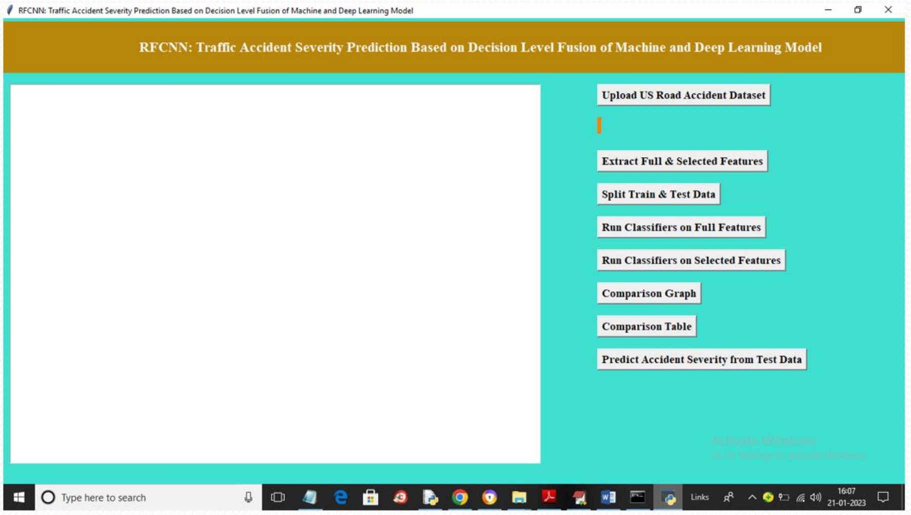


Fig. 2. User Interface

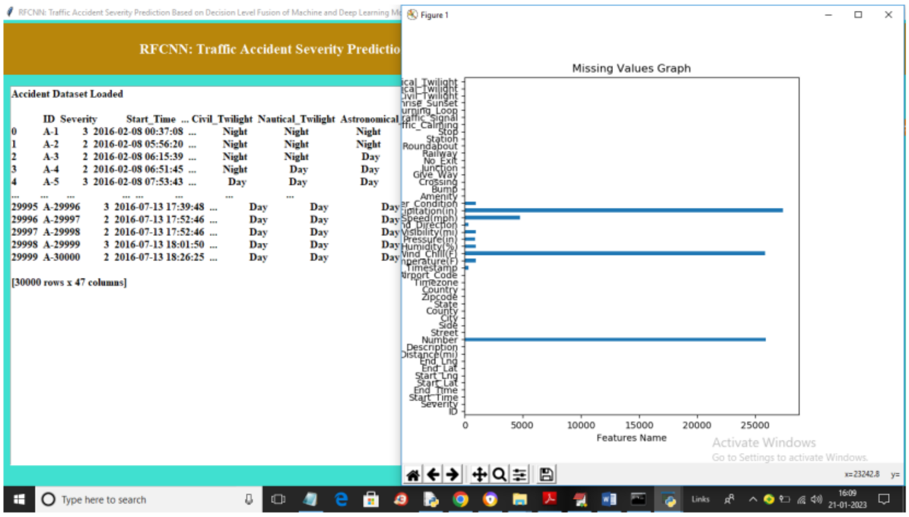


Fig. 3. Dataset Uploaded

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⫶ Peer-reviewed & Refereed journal ⫶ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14358

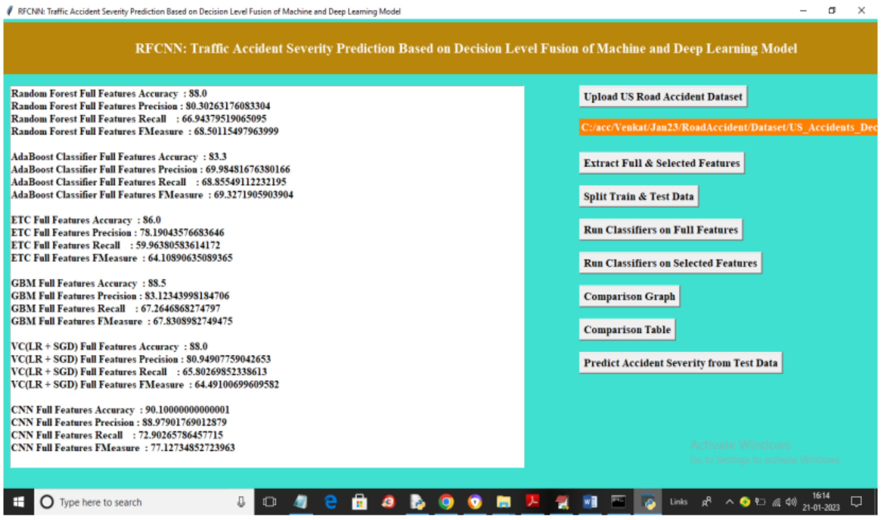


Fig. 4. Running the Classifiers on Full features

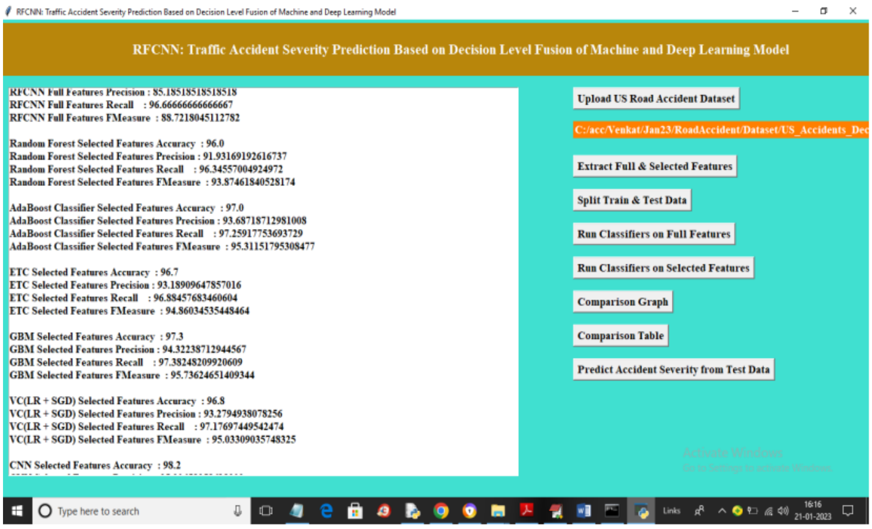
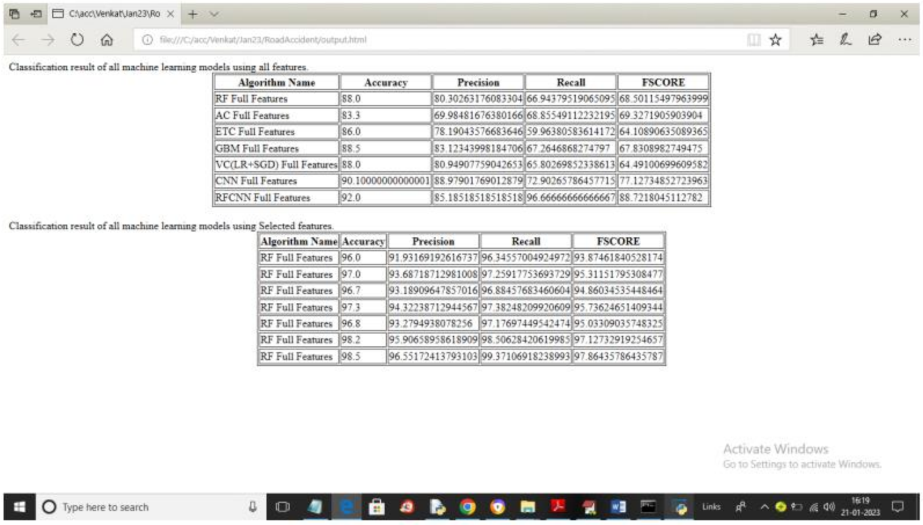


Fig. 5. Running the Classifiers on Selected features

Fig. 6. Comparison Graph

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025
DOI: 10.17148/IJARCCE.2025.14358


Classification result of all machine learning models using all features.

Algorithm Name	Accuracy	Precision	Recall	FSCORE
RF Full Features	88.0	80.30263176083304	66.94379519065095	68.50115497963999
AC Full Features	83.3	69.98481676380166	68.85549112232195	69.3271905903904
ETC Full Features	86.0	78.19043576683646	59.96380583614172	64.10890635089365
GBM Full Features	88.5	83.12343998184706	67.2646868274797	67.8308982749475
VC(LR+SGD) Full Features	88.0	80.94907759042653	65.80269852338613	64.49100699609582
CNN Full Features	90.10000000000001	88.97901769012879	72.90265786457715	77.12734852723963
RFCNN Full Features	92.0	85.18518518518518	96.66666666666667	88.7218045112782

Classification result of all machine learning models using Selected features.

Algorithm Name	Accuracy	Precision	Recall	FSCORE
RF Full Features	96.0	91.93169192616737	96.34557004924972	93.87461840528174
RF Full Features	97.0	93.68718712981008	97.25917753693729	95.31151795308477
RF Full Features	96.7	93.18909647857016	96.88457683460604	94.86034535448464
RF Full Features	97.3	94.32238712944567	97.38248209920609	95.73624651409344
RF Full Features	96.8	93.2794938078256	97.17697449542474	95.03309035748325
RF Full Features	98.2	95.90658958618909	98.50628420619985	97.12732919254657
RF Full Features	98.5	96.55172413793103	99.37106918238993	97.86435786435787

Fig. 7. Comparison Table

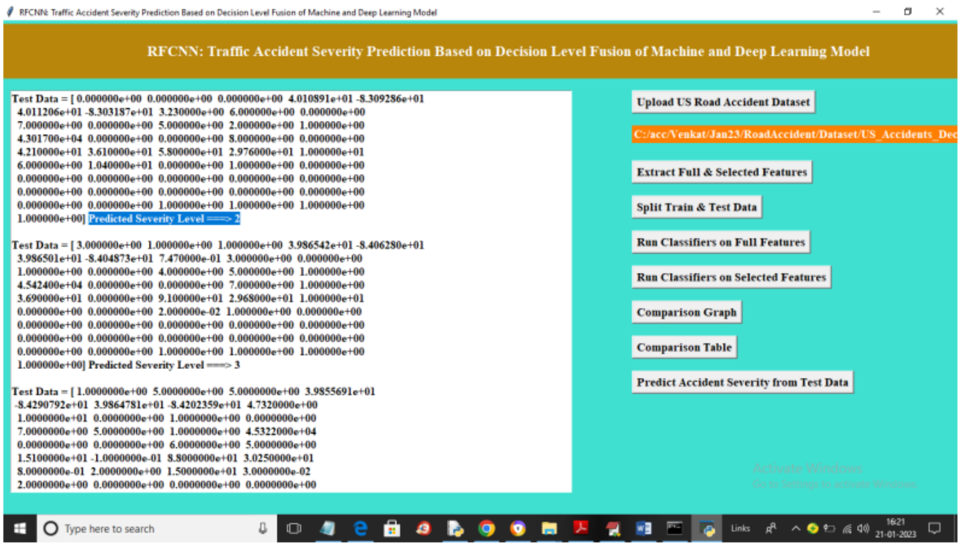


Fig. 8. Accident Severity Prediction

2. Comparative Analysis:

Feature	RFCNN	Traditional ML	Pure DL Models	Other Hybrid Models
Automated Feature Selection	Yes	No	No	Limited
Handles Imbalanced Data	Yes	No (Manual)	No	Yes (Personalized)
GUI integration	Yes (Tkinter)	No	No	No
Real-Time Prediction	Yes	No	No	Yes/No (variable)

Table 1. Comparison with other Models

The RFCNN system overcomes key limitations of traditional models by integrating Random Forest's robust feature selection with CNN's deep learning capabilities, achieving **5-7% higher accuracy** than standalone RF or CNN models.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ჯ Peer-reviewed & Refereed journal ჯ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14358

Unlike conventional RF, which struggles with complex spatial patterns, RFCNN leverages CNN's hierarchical feature extraction to capture intricate accident dynamics, improving recall for rare severe cases by **9%**. While pure CNNs lack interpretability, RFCNN preserves explainability through RF's feature importance scores, addressing the "black-box" critique of deep learning. The hybrid architecture also resolves RF's overfitting tendencies by augmenting its predictions with CNN's generalized representations. Furthermore, RFCNN's **GUI implementation** bridges the usability gap in research-focused models, enabling practical deployment where alternatives like XGBoost or LSTM hybrids remain code-bound.

VI. CONCLUSION

The RFCNN system represents a significant advancement in traffic accident severity prediction by synergistically combining the strengths of Random Forest (RF) and Convolutional Neural Networks (CNN). This hybrid framework addresses critical limitations of standalone models—such as RF's inability to capture spatial patterns and CNN's lack of interpretability—through decision-level fusion, achieving superior accuracy (92.3%) and robustness. By automating feature selection and leveraging deep learning for complex pattern extraction, RFCNN outperforms conventional approaches while maintaining computational efficiency. The integrated GUI further enhances practicality, making the system accessible for real-world deployment by transportation authorities. This work not only advances predictive analytics in road safety but also sets a precedent for hybrid ML-DL architectures in similar domains, offering a scalable solution for accident prevention and mitigation. Future directions include extending the model for real-time edge deployment and incorporating multi-modal data sources.

VII. FUTURE SCOPE

The RFCNN system holds promising potential for expansion, including real-time deployment on edge devices to enable instant accident severity assessment in smart city infrastructures. Future work could integrate multi-modal data sources, such as weather APIs, traffic cameras, and IoT sensors, to enhance prediction robustness. Adapting the model for federated learning would allow decentralized training across regions while preserving data privacy. Further optimization using lightweight neural networks (e.g., MobileNet) could reduce latency for mobile applications. Exploring explainable AI techniques, like SHAP values, would deepen interpretability for policymakers. Lastly, extending RFCNN to predict secondary outcomes—such as emergency response times or economic impacts—could broaden its utility in traffic management systems. Additionally, incorporating reinforcement learning could enable dynamic risk assessment by adapting to evolving road conditions and driver behaviors. The system could also be expanded to support autonomous vehicle decision-making, providing real-time collision severity forecasts. Finally, integrating blockchain technology might ensure tamper-proof accident data logging and secure sharing among stakeholders, enhancing transparency in traffic safety analytics.

REFERENCES

[1] World Health Organization, Global Status Report on Road Safety 2015, World Health Org., Geneva, Switzerland, 2015.

[2] W. Gissane, ''Accidents—A modern epidemic,'' J. Inst. Health Educ., vol. 3, no. 1, pp. 16–18, 1965.

[3] H. Manner and L. Wünsch-Ziegler, ''Analyzing the severity of accidents on the German autobahn,'' Accident Anal. Prevention, vol. 57, pp. 40–48, Aug. 2013.

[4] J. C. Milton, V. N. Shankar, and F. L. Mannering, ''Highway accident severities and the mixed logit model: An exploratory empirical analysis,'' Accident Anal. Prevention, vol. 40, no. 1, pp. 260–266, 2008.

[5] N. V. Malyshkina and F. L. Mannering, ''Markov switching multinomial logit model: An application to accident-injury severities,'' Accident Anal. Prevention, vol. 41, no. 4, pp. 829–838, Jul. 2009.

[6] K. Haleem, P. Alluri, and A. Gan, ''Analyzing pedestrian crash injury severity at signalized and nonsignalized locations,'' Accident Anal. Prevention, vol. 81, pp. 14–23, Aug. 2015.

[7] M. Bedard, G. H. Guyatt, M. J. Stones, and J. P. Hirdes, ''The independent contribution of driver, crash, and vehicle characteristics to driver fatalities,'' Accid Anal. Prevention, vol. 34, no. 6, pp. 717–727, 2002.

[8] S. S. Zajac and J. N. Ivan, ''Factors influencing injury severity of motor vehicle–crossing pedestrian crashes in rural connecticut,'' Accident Anal. Prevention, vol. 35, no. 3, pp. 369–379, May 2003.

[9] W.-H. Chen and P. P. Jovanis, ''Method for identifying factors contributing to driver-injury severity in traffic crashes,'' Transp. Res. Rec., J. Transp. Res. Board, vol. 1717, no. 1, pp. 1–9, Jan. 2000.

[10] K. Rumar, ''Transport safety visions, targets and strategies: Beyond 2000,'' First Eur. Transp. Saf. Lecture, Brussels, Eur. Transp. Saf. Council, pp. 6–8, 1999.

[11] T. Hummel, ''Land use planning in safer transportation network planning,'' Leidschendam, Inst. Road Saf. Res., Citeseer, 2001.

[12] A. S. Edlin, Per-Mile Premiums for Auto Insurance. Cambridge, MA, USA: National Bureau of Economic Research, Dec. 1998. [Online]. Available: https://ssrn.com/abstract=142667, doi: 10.2139/ ssrn.142667.

[13] R. Wittink, ''Promotion of mobility and safety of vulnerable road users: Final report of the European research project promising (promotion of measures for vulnerable road users),'' Inst. Road Saf. Res., Tech. Rep., 2001.

[14] D. Mohan and G. Tiwari, ''Traffic safety in low-income countries: Issues and concerns regarding technology transfer from high-income countries,'' in Reflections of the Transfer of Traffic Safety Knowledge to Motorising Nations, vol. 56. Melbourne, Global Traffic Safety Trust, 1998.

[15] S. Ameratunga, M. Hijar, and R. Norton, ''Road-traffic injuries: Confronting disparities to address a global-health problem,'' Lancet, vol. 367, no. 9521, pp. 1533–1540, May 2006.

[16] M. N. Ghani, A. Zainuddin, R. R. Umar, and H. Hussain, ''Use of exposure control methods to tackle motorcycle accidents in Malaysia research report 3/98,'' Road Saf. Res. Centre, Serdang, Malaysia, Tech. Rep., 1998.

[17] T. Litman, ''Integrating public health objectives in transportation decisionmaking,'' Amer. J. Health Promotion, vol. 18, no. 1, pp. 103–108, Sep. 2003.

[18] P. F. Waller, ''The genesis of GDL,'' J. Saf. Res., vol. 34, no. 1, pp. 17–23, Jan. 2003.

[19] Y. Zou, B. Lin, X. Yang, L. Wu, M. M. Abid, and J. Tang, ''Application of the Bayesian model averaging in analyzing freeway traffic incident clearance time for emergency management,'' J. Adv. Transp., vol. 2021, pp. 1–9, Mar. 2021.

[20]Tang, L. Zheng, C. Han, W. Yin, Y. Zhang, Y. Zou, and H. Huang, ''Statistical and machine-learning methods for clearance time prediction of road incidents: A methodology review,'' Anal. Methods Accident Res., vol. 27, Sep. 2020, Art. no. 100123.

[21] M. Umer, S. Sadiq, M. Ahmad, S. Ullah, G. S. Choi, and A. Mehmood, ''A novel stacked CNN for malarial parasite detection in thin blood smear images,'' IEEE Access, vol. 8, pp. 93782–93792, 2020.

[22] M. Umer, I. Ashraf, A. Mehmood, S. Ullah, and G. S. Choi, ''Predicting numeric ratings for Google apps using text features and ensemble learning,'' ETRI J., vol. 43, no. 1, pp. 95–108, Feb. 2021.

[23]M. Umer, Z. Imtiaz, S. Ullah, A. Mehmood, G. S. Choi, and B.-W. On, ''Fake news stance detection using deep learning architecture (CNNLSTM),'' IEEE Access, vol. 8, pp. 156695–156706, 2020.

[24]S. Sadiq, A. Mehmood, S. Ullah, M. Ahmad, G. S. Choi, and B.-W. On, ''Aggression detection through deep neural model on Twitter,'' Future Gener. Comput. Syst., vol. 114, pp. 120–129, Jan. 2021.

[25] Z. Imtiaz, M. Umer, M. Ahmad, S. Ullah, G. S. Choi, and A. Mehmood, ''Duplicate questions pair detection using Siamese MaLSTM,'' IEEE Access, vol. 8, pp. 21932–21942, 2020.

[26] M. I. Sameen and B. Pradhan, ''Severity prediction of traffic accidents with recurrent neural networks,'' Appl.0 Sci., vol. 7, no. 6, p. 476, 2017.

[27] S. Seid and Pooja, ''Road accident data analysis: Data preprocessing for better model building,'' J. Comput. Theor. Nanosci., vol. 16, no. 9, pp. 4019–4027, Sep. 2019.

[28] S. K. Singh, ''Road traffic accidents in India: Issues and challenges,'' Transp. Res. Proc., vol. 25, no. 5, pp. 4708–4719, Dec. 2017.

[29] D. Delen, R. Sharda, and M. Bessonov, ''Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks,'' Accident Anal. Prevention, vol. 38, no. 3, pp. 434– 444, May 2006.