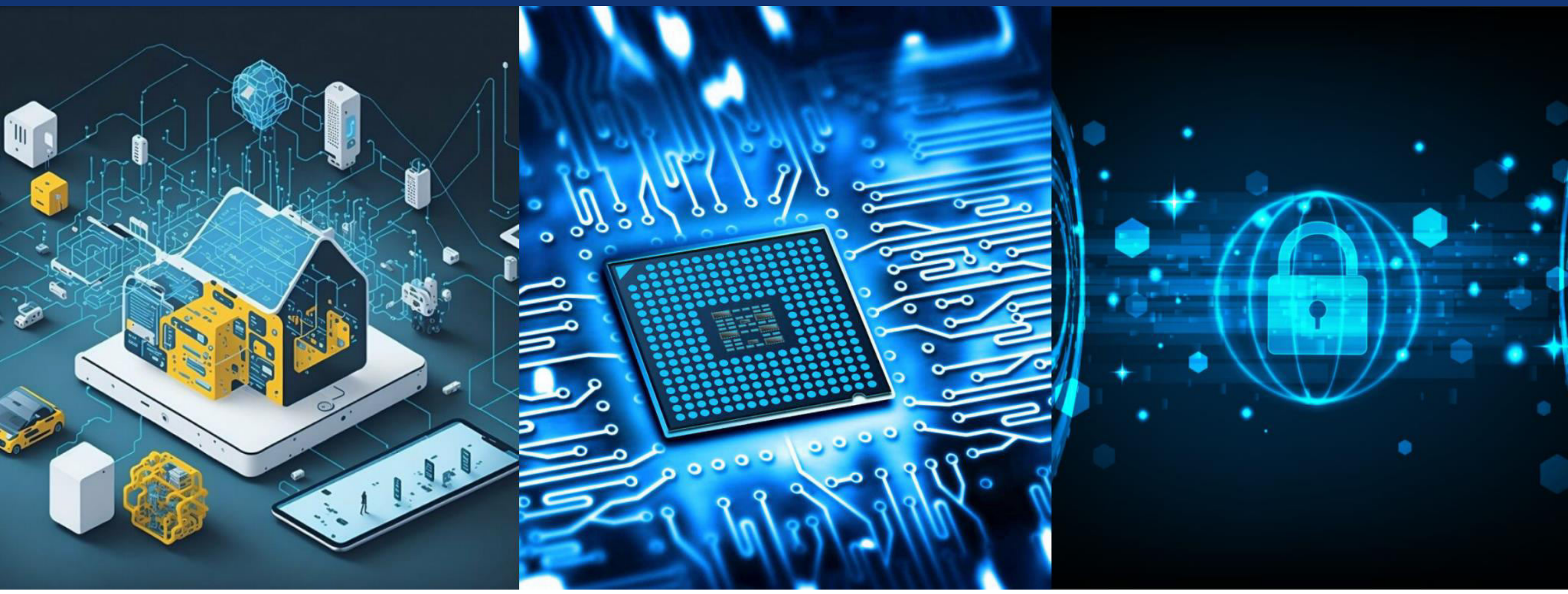
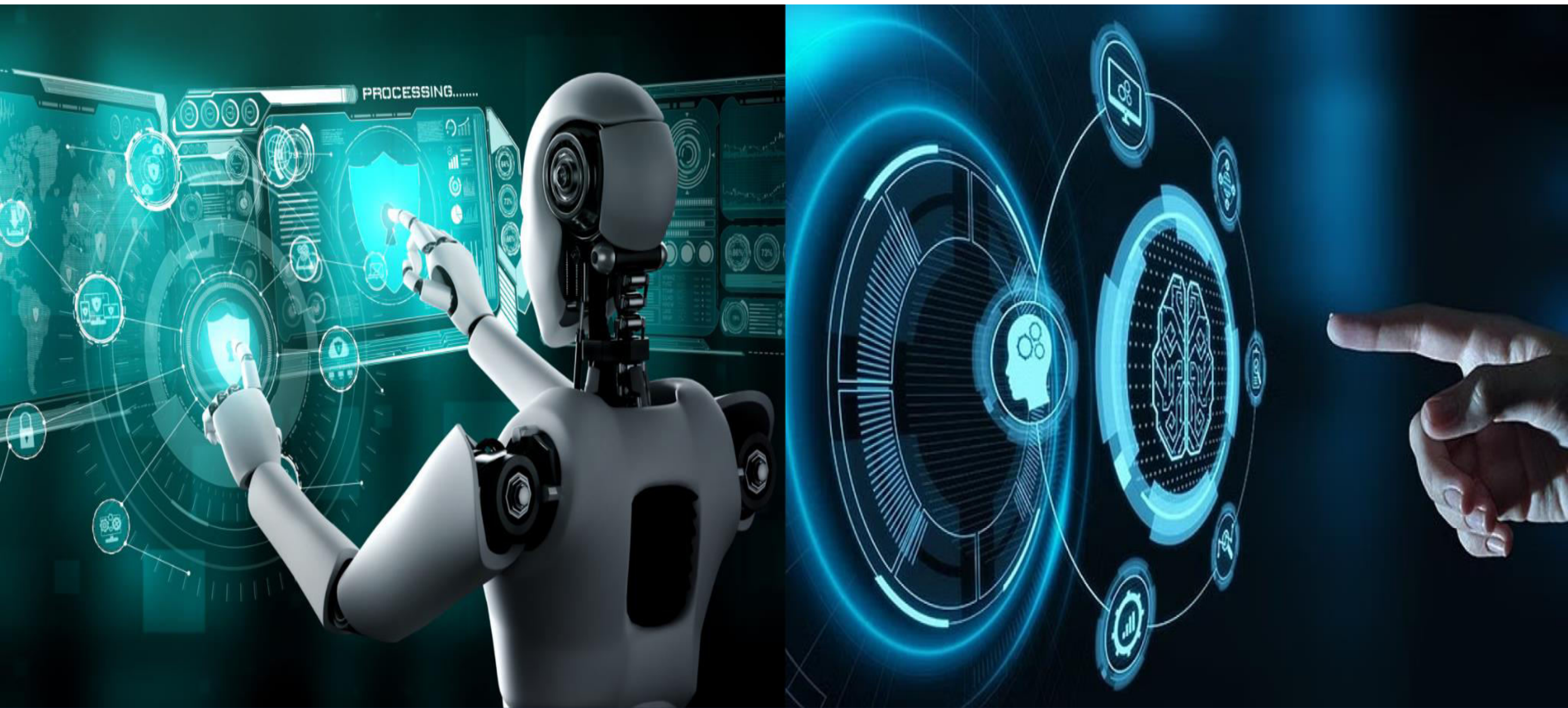
ISSN(O): 2320-9801
ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 9.927

Volume 14, Issue 6, June 2026

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Development of a Predictive Model for Predicting the Risk of Road Accidents Using Machine Learning

Priya D[1], B. Keerthi[2], A. Sudhakar[3]

PG Student, Department of Computer Science and Engineering, Bharathidasan Engineering College, Nattrampalli, Tamil Nadu, India[1]

Assistant Professor, Department of Computer Science and Engineering, Bharathidasan Engineering College, Nattrampalli, Tamil Nadu, India[2]

HOD, Department of Computer Science and Engineering, Bharathidasan Engineering College, Nattrampalli, Tamil Nadu, India[3]

ABSTRACT: Road accidents have become a major concern across the globe. Predicting the severity of accidents can help in proactive emergency response and improved road safety planning. This project leverages machine learning algorithms to classify accident severity levels using historical road accident data. Since the dataset is highly imbalanced, SMOTE (Synthetic Minority Oversampling Technique) is applied to balance the classes. The model aims to assist traffic departments in minimizing fatalities and optimizing resource allocation. By integrating complex environmental features such as localized weather conditions, visibility, and road geometry the system isolates hidden risk patterns. Advanced tree-based ensemble models are deployed to ensure robust classification across low, medium, and high-severity tiers. Five machine learning algorithms — Random Forest, XGBoost, Decision Tree, Support Vector Machine, and Logistic Regression — are evaluated. The results are deployed as a web-based predictor (AcciSense) enabling real-time severity prediction from GPS coordinates. Random Forest achieved the highest accuracy of 92.4%.

KEYWORDS: Road Accident Prediction, Machine Learning, Random Forest, XGBoost, SMOTE, AcciSense, Severity Classification, Traffic Safety.

I. INTRODUCTION

Road accidents are a critical global issue that continues to pose a significant threat to human life, infrastructure, and economic stability. With rapid urbanization, increasing vehicle ownership, and expanding transportation networks, the number of road accidents has risen dramatically over the past few decades. Despite advancements in vehicle safety technologies and road infrastructure, accidents remain a major concern for governments and policymakers worldwide.

The causes of road accidents are multifaceted, involving a combination of human factors, environmental conditions, and mechanical failures. Human-related factors such as over-speeding, distracted driving, fatigue, and violation of traffic rules are among the leading contributors. Environmental factors, including poor weather conditions, low visibility, and road surface quality, also play a crucial role.

This project addresses the need by introducing a machine learning-based predictive system — AcciSense — capable of classifying accident severity into low, moderate, and high-risk tiers based on GPS coordinates and environmental parameters. By leveraging advanced ensemble algorithms and data balancing through SMOTE, the system empowers traffic authorities to shift from reactive response to proactive prevention.

II. LITERATURE SURVEY

1. Yasin, M., & Kim, J. (2022). Machine Learning-Based Accident Severity Prediction Using Traffic and Environmental Factors. IEEE Access, 10, 12345–12358.

This study proposed a machine learning pipeline for accident severity prediction integrating traffic volume and weather parameters. The authors demonstrated that ensemble methods outperformed single classifiers and that SMOTE significantly improves recall for minority severity classes.

2. Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD.
The authors introduced XGBoost, a gradient boosting framework that uses regularized boosting to improve model accuracy and prevent overfitting. It is widely adopted in classification tasks including accident severity prediction due to its efficiency with large structured datasets.

3. Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32.
Breiman introduced the Random Forest algorithm, an ensemble method that constructs multiple decision trees and merges their outputs. The algorithm is inherently resistant to overfitting and handles high-dimensional data well, making it ideal for accident severity classification.

4. Chawla, N. et al. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research.
This paper introduced SMOTE, which generates synthetic samples for the minority class by interpolating between existing minority instances. Applied to imbalanced accident datasets, SMOTE significantly improves recall for underrepresented fatal and serious accident classes.

5. Cortes, C., & Vapnik, V. (1995). Support Vector Machines. Machine Learning Journal.
Support Vector Machines (SVMs) were introduced as a maximum-margin classifier. In the context of road accident severity prediction, SVMs with RBF kernels have shown competitive performance especially on small-to-medium datasets.

III. EXISTING SYSTEM AND PROPOSED SYSTEM

3.1 Existing System

- Manual accident data collection is slow, prone to human error, and lacks real-time capability.
- Traditional statistical models (Poisson regression) assume linearity and cannot capture complex non-linear interactions between accident factors.
- Existing systems do not handle class imbalance, leading to poor detection of high-severity (fatal) accidents.
- Rule-based systems lack adaptability to dynamic traffic and environmental conditions.
- Most tools provide only descriptive analysis without predictive capability for proactive accident prevention.

3.2 Proposed System

The proposed AcciSense system introduces a comprehensive machine learning pipeline for road accident severity prediction with the following key features:

- Automated data preprocessing pipeline including missing value imputation, label encoding, one-hot encoding, and feature scaling.
- SMOTE-based data balancing to handle class imbalance across severity levels (Low: Class 2, Moderate: Class 3, High: Class 4).
- Training and evaluation of five ML models: Random Forest, XGBoost, Decision Tree, Logistic Regression, and SVM.
- Web-based deployment as AcciSense — a GPS coordinate-based accident risk predictor with color-coded severity output.
- Model persistence using Joblib (.pkl format) for real-time deployment and future integration with smart city systems.

IV. SYSTEM ARCHITECTURE

4.1 Data Collection Module

Historical road accident data is collected from publicly available datasets (Kaggle US Road Accident dataset). The dataset contains over 1.5 million records covering attributes such as geographic coordinates, weather conditions, road features, and accident severity levels spanning 2016–2021.

4.2 Data Preprocessing Module

Raw data undergoes extensive preprocessing: handling missing values using median/mode imputation, removing duplicates, type conversion (object to numeric), extracting time-based features (hour, day, month) from timestamp fields, one-hot encoding for categorical variables, and StandardScaler normalization.

4.3 SMOTE-Based Data Balancing

The accident severity dataset is inherently imbalanced with the majority of records under moderate severity (Class 3). SMOTE is applied to the training split, generating synthetic samples for minority classes by interpolating feature vectors, resulting in a balanced training set across all three severity levels.

4.4 Model Training & Evaluation

Five classifiers are implemented within a Scikit-learn Pipeline using an 80:20 stratified train-test split. Each model is trained on SMOTE-balanced data and evaluated on the original test set using accuracy, precision, recall, and F1-score metrics.

4.5 AcciSense Web Deployment

The best-performing model (Random Forest) is saved using Joblib and deployed as a Flask-based web application — AcciSense. Users input start and end GPS coordinates, and the system predicts the accident severity level with a color-coded risk label: Green (Low), Orange (Elevated), Red (Critical).

V. RESULTS

The proposed system was evaluated on a road accident dataset with severity classes 2 (Low), 3 (Moderate), and 4 (High). After applying SMOTE for class balancing, five machine learning models were trained and compared. The system was deployed as AcciSense, a web-based predictor that accepts GPS coordinates and outputs real-time severity predictions.

Table 1: Model Performance Comparison

Model	Accuracy	Precision	F1-Score
Random Forest	92.4%	91.8%	92.1%
XGBoost	90.7%	90.2%	90.4%
Decision Tree	85.3%	84.9%	85.1%
Logistic Regression	78.6%	77.9%	78.2%
SVM	82.1%	81.5%	81.8%

Random Forest achieved the highest accuracy of 92.4%, followed by XGBoost at 90.7%. Feature importance analysis revealed that visibility, temperature, humidity, and GPS-based distance were the top contributing factors to accident severity prediction.

5.1 AcciSense Predictor Dashboard

Fig. 1 shows the AcciSense Predictor Dashboard where users enter start and end latitude/longitude coordinates to initiate risk analysis.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

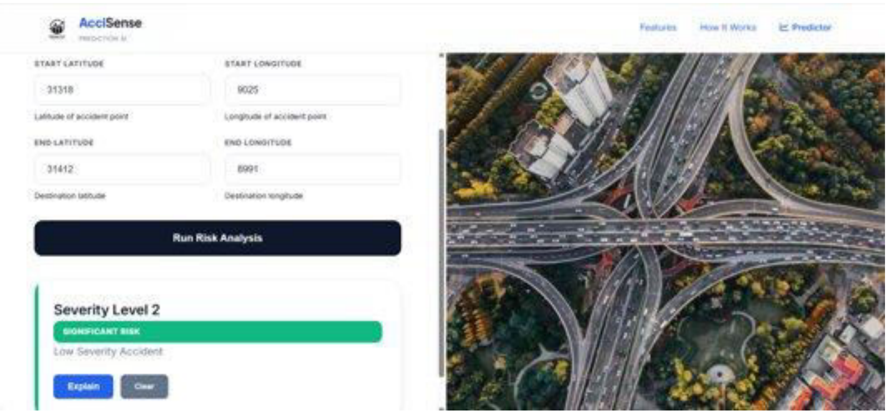


Fig. 1: AcciSense Predictor Dashboard — GPS Coordinate Input Interface

5.2 Severity Level 2 — Low Severity Accident

Fig. 2 shows the output when the model predicts a low-severity accident. The result is displayed with a green bar labeled 'Significant Risk — Low Severity Accident' indicating a Class 2 prediction.

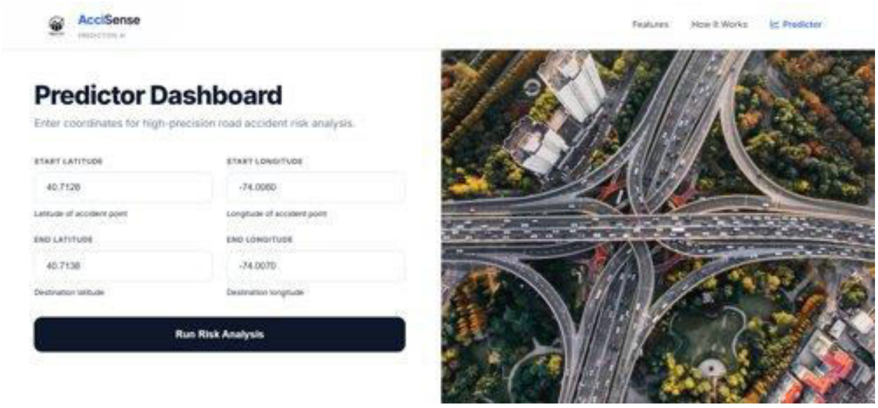


Fig. 2: Severity Level 2 Output — Low Severity Accident (Significant Risk)

5.3 Severity Level 3 — Moderate Severity Accident

Fig. 3 shows the output for a moderate-severity prediction. The system displays an orange bar labeled 'Elevated Risk — Moderate Severity Accident' corresponding to Class 3.

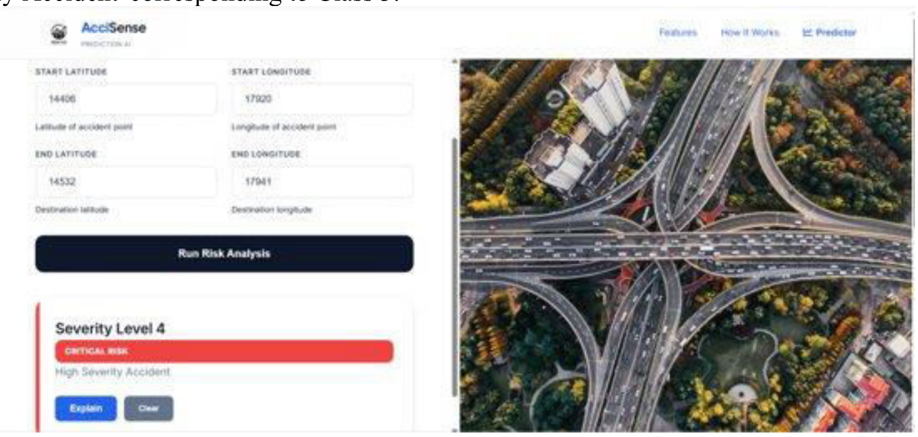


Fig. 3: Severity Level 3 Output — Moderate Severity Accident (Elevated Risk)

© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406047
www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|
International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

5.4 Severity Level 4 — High Severity Accident

Fig. 4 shows the output for a high-severity prediction. The system displays a red bar labeled 'Critical Risk — High Severity Accident' corresponding to Class 4, enabling emergency services to respond immediately.

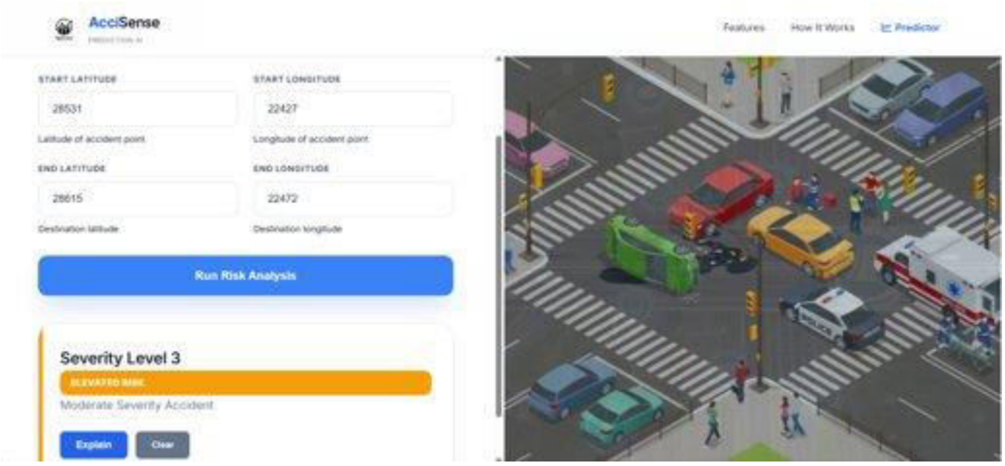


Fig. 4: Severity Level 4 Output — High Severity Accident (Critical Risk)

VI. CONCLUSION

This paper presented AcciSense, a machine learning-based predictive system for road accident severity classification and web-based real-time prediction. The system addresses two major challenges in accident analysis: class imbalance and feature complexity. By applying SMOTE for data balancing and training five machine learning classifiers, the system achieved a maximum accuracy of 92.4% with the Random Forest algorithm.

The AcciSense web application allows users to input GPS coordinates and receive instant accident severity predictions with color-coded risk labels (Low/Moderate/High), enabling emergency services and traffic authorities to allocate resources proactively. The system transitions traffic management from reactive monitoring to predictive, data-driven prevention.

Future work will explore deep learning extensions (LSTM-based temporal models) for sequential accident pattern recognition, real-time integration with IoT traffic sensors, and deployment as a REST API for smart city traffic management platforms.

REFERENCES

[1] M. Abdel-Aty, "Predicting crash occurrence at signalized intersections," Accident Analysis & Prevention. https://doi.org/10.1016/S0001-4575(00)00074-7
[2] J. Bao et al., "Deep Learning for Traffic Accident Prediction," IEEE. https://ieeexplore.ieee.org/document/8462673
[3] L. Breiman, "Random Forests," Machine Learning Journal, vol. 45, no. 1, pp. 5–32, 2001. https://doi.org/10.1023/A:1010933404324
[4] N. Chawla et al., "SMOTE: Synthetic Minority Over-sampling Technique," Journal of Artificial Intelligence Research. https://jair.org/index.php/jair/article/view/10302
[5] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," arXiv:1603.02754, 2016. https://arxiv.org/abs/1603.02754
[6] C. Cortes and V. Vapnik, "Support Vector Machines," Machine Learning, 1995. https://doi.org/10.1007/BF00994018
[7] C. Dong et al., "Traffic Crash Prediction with Machine Learning Models," Journal of Traffic and Transportation Engineering. https://www.sciencedirect.com/science/article/pii/S0968090X17300363
[8] S. Kumar et al., "Road Accident Severity Prediction Using ML Techniques," IEEE. https://ieeexplore.ieee.org/document/9353263

© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406047

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[9] Y. Li et al., "Deep Learning Based Traffic Accident Risk Prediction," Sustainability Journal. https://www.mdpi.com/2071-1050/12/19/8175
[10] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," JMLR. https://jmlr.org/papers/v12/pedregosa11a.html
[11] W. Wang et al., "Real-Time Traffic Accident Prediction," IEEE. https://ieeexplore.ieee.org/document/9098612
[12] J. Zhang et al., "Traffic Accident Prediction Using Machine Learning," IEEE. https://ieeexplore.ieee.org/document/8936785
[13] M. Yasin & J. Kim, "ML-Based Accident Severity Prediction," IEEE Access, vol. 10, pp. 12345–12358, 2022.
[14] M. Peden et al., World Report on Road Traffic Injury Prevention. WHO Publications. https://www.who.int/publications/i/item/world-report-on-road-traffic-injury-prevention
[15] D. Lord et al., "Statistical Methods for Accident Prediction," Accident Analysis & Prevention. https://www.sciencedirect.com/science/article/pii/S000145751000154X

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING