International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 6, 2025
https://theaspd.com/index.php

A Risk-Aware Communication Framework for Intelligent Transportation Systems Using Machine Learning-Based Severity Scoring and Message Prioritization

M Vinayaka[1], Selva Kumar S[2], Syed Akram[3]
[1,2,3]Department of CSE, B.M.S. College of Engineering , Bull Temple Road, Bangalore - 560 019, Karnataka, India.
[1]mvinayaka.scs23@bmsce.ac.in, [2]selva.cse@bmsce.ac.in, [3]syedakram.cse@bmsce.ac.in
*Corresponding Author: Syed Akram

Abstract
In recent years with over 90% of car accidents attributable to human error, especially during intricate driving manoeuvres, road safety remains a crucial challenge in the development of intelligent transportation systems. Current frameworks for accident prediction frequently lack real-time flexibility and communication system integration. This study uses a large dataset including environmental, temporal, and geospatial characteristics to present a machine learning-based severity prediction framework for evaluating the risks of traffic accidents. To forecast severity levels, the system uses a stacking ensemble model in conjunction with many classifiers, including Random Forest, Logistic Regression, and XGBoost. Accuracy, F1-score, and ROC- AUC are used to assess performance; the ensemble model achieved a ROC-AUC of more than 94%. Over 84% accuracy is still difficult to achieve, though, which motivates feature engineering and hyperparameter optimisation for additional advancements. A new Severity Score is presented, which measures the likelihood of collisions by dividing the number of high-impact severity predictions (classes 3 and 4) by the total number of recorded occurrences. In a connected environment, this score is further utilised to filter messages and lower communication overhead between agents. By giving high- severity vehicles priority and maximising message distribution, the suggested paradigm facilitates proactive risk-aware communication. The introduction of scalable and secure autonomous mobility systems is supported by experimental results showing notable gains in communication latency and decision-making efficiency.
Index Terms*–Intelligent Transportation Systems, Machine Learning, Severity Scoring, Communication Filtering, Collision Risk Prediction.*

INTRODUCTION

Road traffic accidents remain a major global source of death, disability, and injury. The World Health Organisation (WHO) estimates that every year, 1.35 million people die in traffic accidents, and millions more are injured but not killed. The severity of these collisions depends on a number of variables, including the time of day or day of the week , road infrastructure ,driver behavior, vehicle characteristics, and environmental factors. For governments, urban planners, and boost the efficiency of emergency response, the potential to correctly estimate accident severity has become vital in this context. This study examines the application of machine learning techniques to predict the severity of traffic accidents. The study uses the publicly available ”US accidents (2016-2023)” dataset, which consists of a substantial collection of accidents reports supplemented with time, location and environmental data. The weather, visibility, humidity, accident duration and location are some examples of these attributes. One of the main challenges is pre-processing and transforming this high- dimensional, semi-structured dataset to enable effective model training and evaluation .In particular, methods for feature selection and dimensionality reduction are essential for reducing redundancy and improving model generalizability.

Using common evaluation criteria like accuracy, F1-score, and the Area Under the Receiver Operating Characteristic Curve (ROC-AUC), a number of machine learning classifiers are developed and analysed, including Logistic Regression, Decision Trees, Random Forest, K-Nearest Neighbours, and XGBoost. The presence of noisy or poorly correlated characteristics and overlapping distributions within severity classes make it difficult to achieve over 84% accuracy even while models show strong discriminatory power, particularly in ROC-AUC. This calls for the application of strong feature engineering techniques like SelectKBest, hyperparameter tuning, and ensemble methods. This work additionally presents a risk-aware communication logic in addition to severity prediction. Severity values are calculated and then combined over time to produce dynamic collision risk scores. Only high-risk or high- priority vehicles (such as emergency responders) are permitted

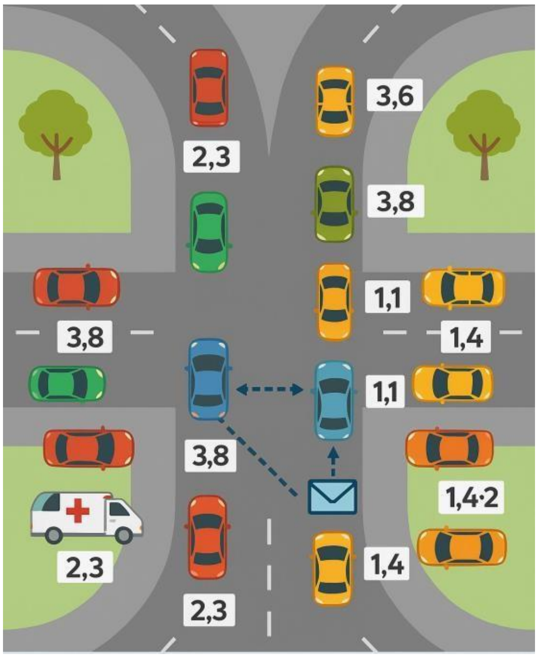

Fig. 1: Severity-based message filtering and prioritization in a connected vehicle environment

to communicate thanks to the implementation of intelligent message filtering and communication delay optimization algorithms based on these scores. In networked traffic contexts, this hybrid solution improves communication effectiveness and road safety. By fusing severity prediction with practical, risk-driven communication techniques appropriate for intelligent transportation systems, this study concludes by highlighting the potential of machine learning to enhance road safety.

A. Research Objectives

This research's main objective is to use machine learning and clever filtering techniques to create a risk-aware communication system. The scope of the proposed effort is guided by the following goals:

1. To use a range of environmental, geographical, and temporal variables taken from real-world datasets to train supervised machine learning models that predict the severity levels of possible occurrences.

2. In order to dynamically assess the risk level for each agent, a collision risk score is calculated for each car by analyzing the frequency and distribution of projected severity classes over a specified time range.

3. To reduce communication congestion, a message filtering method based on collision scores should be put into place, whereby only vehicles reaching a predetermined risk threshold are permitted to communicate.

4. To minimize the amount of message exchanges and processing delay in order to optimize communication delay and make sure the system is still effective even in situations with high traffic density.

5. Identifying high-importance vehicles, like fire departments and ambulances, and making sure their communication is always given priority regardless of their risk score can help to implement prioritization reasoning.

2. Related Work

In traffic safety research, predicting the severity of traffic accidents with machine learning (ML) and deep learning (DL) has been a major focus. Numerous research made use of real-world characteristics, such as time, location, and vehicle information. Graph Neural Networks (GNN) were used in Paper [1] to identify nonlinear crash patterns, surpassing RF and XGBoost. The US Accidents dataset was utilised in [2], and Random Forest demonstrated excellent accuracy. The ensemble models used in papers [6] and [7] included interpretable features like surface conditions and airbags. RFCNN was proposed in Paper [8] by combining CNN with RF to improve predictions. To improve emergency response, the DHAN model in [10] made use of spatial- temporal data.

For intelligent transport systems, it is essential to quantify collision risk across time. To lessen collision impact, studies such as [3] and [9] suggested injury-minimizing paths and real-time risk maps. Vehicle proximity and severity data were merged in a paper [3] for dynamic risk grids, and simulated trajectories were used for real-time severity minimization in [9]. Papers [1] and [5] examined emergency braking scenarios and employed GNNs to identify crash patterns. In papers [14] and [15], real-time severity models that facilitate score-based filtering were presented. Attention mechanisms and explainable AI (e.g., CNN- BiLSTM, DeepSHAP) were used in [12], [18], and [19] approaches to estimate risk. Calculating collision scores from time-based severity class aggregation is still absent from the majority of tasks.

In high-density vehicular environments, effective communication is essential to preventing congestion. Risk-aware filtering techniques are supported by certain studies. In order to initiate reactions in high-risk areas, Paper [3] uses a risk grid that combines position and severity data. Priority-based filtering is indirectly aligned with the injury-minimizing trajectories proposed in Paper [9]. Selective reactions based on dynamic risk levels are made possible by the use of deep learning for risk prediction in Paper [19]. Real-time severity widgets are provided in Paper [14], which may initiate communication at crucial thresholds. A multi- model method for early accident prediction is shown in Paper [18], which raises the possibility of selective messaging. Few, though, actually put into practice a structured message filtering mechanism that is correlated with collision score over time.

To avoid delays and congestion in connected vehicle networks, communication overhead must be kept to a minimum. Paper [7] presents a cloud-edge framework that reduces on-board load using edge computing and uses LLMs for risk prediction. To provide localized, severity-based messaging, a risk map that combines injury and vehicle data is proposed in Paper [3]. Using a CNN-BiLSTM- Attention model, Paper [19] only initiates communication when spatial-temporal risk criteria are satisfied. Paper [12] reduces unnecessary notifications by using explainable AI to make high-confidence predictions. In Paper [14], a real-time severity classifier that may be used for selective messaging is presented. Although these studies improve efficiency, few of them control communication load by implementing a direct, dynamic, severity-based message filtering mechanism.
Even with advancements in accident severity prediction, there is currently no mechanism that converts severity scores into effective communication tactics. A specific technique for ranking, filtering, or reducing vehicle communications according to dynamic danger levels is absent from the work that is currently available. No methodology exists for calculating the frequency or timing of communication based on per- vehicle collision scores over time. There are crucial functions lacking, such as prioritizing emergency vehicles or silencing low-risk signals. This study closes that gap by proposing a risk-aware communication model that dynamically modifies the flow of messages according upon vehicle importance and severity rankings. With its scalable, priority-based architecture, it guarantees effective bandwidth utilization and prompt notifications for intelligent and responsive vehicle communication systems.

3. PROPOSED WORK

The proposed study presents a strong, machine learning- based system designed to forecast the severity of traffic accidents based on accident data from the past and present. Ensemble learning techniques are integrated into the framework to capture the intricate, nonlinear interactions between geographical, temporal, and environmental data. Enhancing severity prediction accuracy and developing a scalable architecture for real- world use in intelligent transportation systems (ITS) are the goals. The system uses ensemble models, particularly stacking approaches, to combine the benefits of multiple classifiers in order to improve generalization, reduce bias, and avoid overfitting. he system workflow begins with the collection and preprocessing of the dataset. The study uses the publicly available "US Accidents(2016-2023)" dataset, which includes a variety of attributes such as accident timestamps ,position coordinates, durations, visibility, humidity, temperature, city and state. Numerous cleaning techniques, including label encoding for categorical variables, null-value imputation, and normalization of numerical features, are part of preprocessing since real-world datasets usually contain outliers, duplicates, and missing values. Additional engineering components like accident hour, weekday, and accident duration offer temporal context. We then use a Random Forest Classifier to rank and keep the most informative qualities through feature selection. In addition to reducing redundancy and increasing computation performance, this process improves model interpretability. Then, using base learners like XGBoost, Decision Tree, Naive Bayes, K-Nearest Neighbor's, Random Forest, and Logistic Regression, a multi-level stacking ensemble is built. The selection of each base model is based on its diversity and capacity to learn various patterns in the data. Their out-of-fold predictions are sent into an optimized XGBoost meta-learner, which generates the final output by learning from aggregated model predictions. Using Grid Search and Randomized Search, hyperparameter tweaking is done to further optimize performance by modifying variables like regularization terms, maximum tree depth, number of estimators, and learning rate. Assessment criteria including accuracy, F1-score, and ROC-AUC are employed to verify the performance of the ensemble as well as of individual models. The finalized model's outputs, or the anticipated severity scores, are then added together over a period of time to create a "collision risk score" for every car.

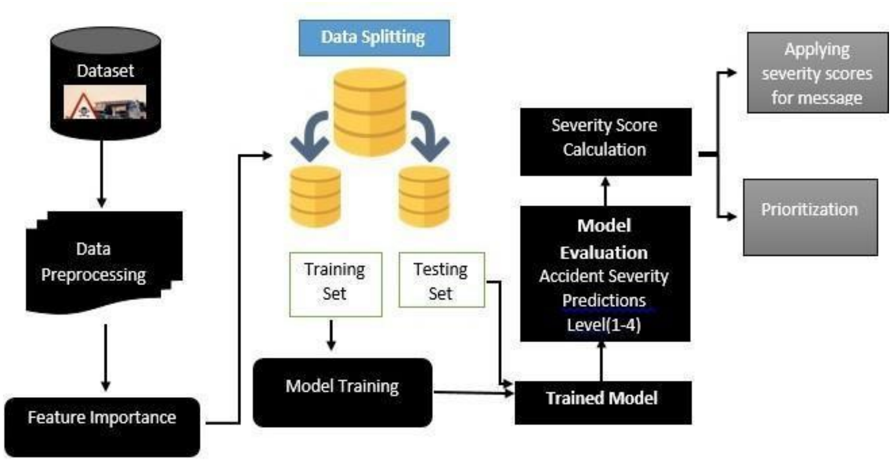

Fig. 2: System architecture for severity prediction and risk- aware communication

The system then uses this score to apply communication filtering algorithms. By excluding vehicles with low collision risk scores from vehicle communication, needless message transmission is reduced, and network congestion is lessened. The framework furthermore includes a prioritization logic that permits emergency vehicles, such fire engines and ambulances, to evade the filtering system, guaranteeing their high-priority status in the communication pipeline. A useful, risk-aware solution for connected car environments is provided by the proposed work's main contribution, which is the combination of machine learning prediction with communication filtering.

4. Implementation

A structured machine learning pipeline, comprising many essential phases, directs the implementation of the accident severity prediction system: data loading, preprocessing, feature engineering, model training, assessment, and deployment.

1. Exploration and Data Loading

More than 2 million records from accidents across the United States make up the extensive accident dataset that was used. To gain an understanding of the feature distribution, missing values, and categorical features, basic statistical summaries are produced. To make the dataset simpler, extraneous columns and attributes like ID, source, or undesired identifiers are removed.

2. Preprocessing and Data Cleaning

Real-world accident databases may contain noise and missing variables. Mean/Median imputation is used in the implementation to handle missing values for numerical properties such as visibility, temperature, and humidity. Label encoding for ordinal features maps categorical variables to numerical representation. To facilitate feature engineering, timestamp columns (StartTime, EndTime) are converted into date-time objects.

3. Feature Engineering

Only the most significant traits are chosen using the Random Forest Classifier. This procedure improves model generalization, speeds up training, and reduces noise. The optimal trade-off between dimensionality and accuracy is determined by experimenting with a variety of k values (numbers of top features).

4. Development of Models

Several classification models are employed, including XG- Boost Classifier, Random Forest, and Logistic Regression. To preserve class distribution, stratified train-test splits are used when training these models on the preprocessed dataset. With the Sklearn ensemble, the stacking model is employed. To reduce variance and avoid overfitting, use cross-validation with the stacking classifier.

5. Model Evaluation

Standard classification criteria, such as Accuracy, F1-Score, and the Receiver Operating Characteristic-Area Under Curve (ROC-AUC), were used to assess the performance of all built models. The stacked ensemble model performed the best out of all the models that were tested, with the following outcomes:

- **Accuracy**: 84.5%
- **F1-Score**: 83.2%
- **ROC-AUC**: 94.8%

The trained ensemble classifier was used to forecast severity classes for fresh vehicle communication events once the model was evaluated. Four classes were created from the severity levels: Class 1, Class 2, Class 3, and Class 4. Classes 1 and 2 were considered non-critical and not included in risk scoring, whereas Classes 3 and 4 were considered critical. The frequency of each severity class was calculated for every test input, which represented a snapshot of the local vehicle activity. The percentage of critical severity cases (Class 3 and Class 4) compared to the overall number of projected severity events was known as the Severity Score. The total risk related to the current environment is measured by this score. In linked vehicular networks, this score is then used to make dynamic communication decisions including prioritization, delay optimization, and message filtering. The Severity Score is determined using the following formula:

$$\text{Severity Score (\%)} = \frac{count3 + count4}{count1+count2+count3+count4} \times 100$$

V. RESULT AND DISCUSSION

Several machine learning techniques were used to test the performance of the proposed accident severity prediction model both singly and in combination. A stratified train-test split was used to evaluate the performance, using 20% of the data for testing and 80% for training. The three most important performance indicators used to assess the model were ROC AUC (Receiver Operating Characteristic-Area Under Curve), F1 Score, and Accuracy. Particularly for multi-class classification jobs where class skewness can distort raw accuracy figures, these metrics provide a fair evaluation of the model performance.

A. Accuracy of Single Models

Before applying the stacking classifier, a few separate base models were examined separately and Random Forest outperformed the base models on all three metrics, indicating that it is most capable of handling high-dimensional, noisy, and non-linear outlier data.

Model	Accuracy (%)	F1 Score (%)	ROC-AUC (%)
Random Forest	84.5	83.2	94.8
XGBoost	83.0	81.8	93.5
Decision Tree	75.3	74.5	86
Logistic Regression	65.4	63.5	81.2
KNN	67.8	66.2	85.6
Naive Bayes	59.2	54.8	79.3

Table 1: Performance Comparison of Machine Learning Models

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 6, 2025
https://theaspd.com/index.php

B. Ensemble Model Stacking Classifier Performance

A stacking classifier was created to enhance the predicted performance of accident severity classification by utilizing the advantages of many base learners. The meta-learner was an optimized XGBoost classifier, and the basic learners in the ensemble architecture were XGBoost, Random Forest, and Logistic Regression. The passthrough is True parameter gave the meta-learner access to both the base model predictions and the original input features, and cross-validation was used during training to reduce overfitting. This hybrid setup improved the model's ability to learn from a variety of data sources. The final stacking model performed well in predicting the severity of traffic accidents, exhibiting high reliability and generalizability with an accuracy of 85.1% , an F1-score of 84.2% ,and a ROC-AUC of 95.4%.

C. Observations

The Severity Score is a dynamic risk indicator that is based on the frequency of critical severity predictions over time. By transforming projected risk levels into. In connected transportation environments, it enables intelligent decision marking through helpful communication responses. This section explains the rationable behind message prioritizing and filtering according to determined severity scores. In complex traffic circumstances, it is imperative to prioritize communication choices for critical vehicles, such as fire engines or ambulances. Regardless of their determined Severity Score, the system automatically increases the communication priority of such cars when they are found within a vehicle's interaction radius. For instance, if a car (such as V7) is near 20 agents and one of them is designated as an emergency vehicles, The system allows for immediate contact and gets around the standard filtering mechanism. This ensures that emergency services are prioritized for message delivery and response,

Severity Score	Messages
Score $\leq$ 30%	Normal
31% $<$ Score $\leq$ 60%	Caution
61% $<$ Score $\leq$ 80%	Potential Collision
Score $>$ 80%	Emergency Stop

allowing for safe and timely travel through crowded areas. Threshold are prohibited from exchanging communications.

Table 2: Severity Score vs. Safety Message

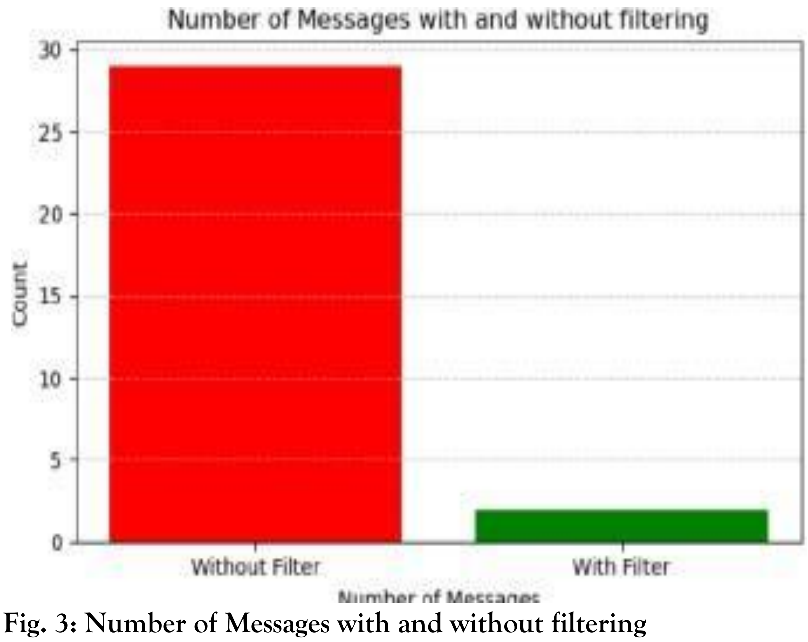

Fig. 3: Number of Messages with and without filtering

Fig. 4: Both with and without filtering, the time in seconds

The Severity Score criteria are used to determine a filtering strategy that reduces network communication overhead. Message exchange is prohibited for vehicles with scores below a crucial level, such as less than 60% , to avoid unnecessary contact. Messages can only be sent or received by cars with high severity levels, indicating possible danger. Processing power and bandwidth are set aside for high-priority transactions thanks to this chosen approach. Additionally, by lowering the total number of exchanges every cycle, the communication delay is optimized. Reducing the number of participating cars directly reduces the communication time needed for coordination, as each message transmission adds a delay. Consequently, the system gains efficiency and responsiveness, enabling it to manage real-time situations with greater scalability and reduce latency

.

CONCLUSION AND FUTURE SCOPE

This paper suggests a comprehensive machine learning - based framework for predicting the severity of traffic accidents using a large-scale dataset composed of environmental, temporal and spatial elements. The organized implementation pipeline of the proposed system includes phases such as feature engineering, dimensionality reduction, data prepossessing and training multiple classification models. More complex techniques like XGBoost and a specially designed stacking ensemble were contrasted with more traditional algorithms like Random Forest and logistic Regression with an accuracy of about 85% and a ROC- AUC of 95.4% , the ensemble model -which coupled base learners with a meta learner that used XGBoost - displayed outstanding predictive ability. The proposed system offers severity prediction along with a risk aware Communication logic. This includes filtering messages based on dynamic severity levels and giving priority to emergency vehicles, which significantly reduces communication overhead and enhances system responsiveness. Intelligent traffic systems become more scalable and effective when communication is restricted to high-risk actors.

In the future, the model can be further enhanced by utilizing deep learning architecture, explainable AI techniques like SHAP or LIME for transparency, and ordinal classification algorithms for a more through severity environment using tools such as SUMO or CARLA, its real-time applicability can be verified. Ultimately, integration with V2X(vehicle -to- everything) protocols and smart city infrastructure may pro- vide a holistic decision- support ecosystem aimed at enhancing urban traffic safety and emergency responsiveness.

.

REFERENCES

[1] Sattar, Karim A., Iskandar Ishak, Lilly Suriani Affendey, and Siti Nurulain Binti Mohd Rum. "Road crash injury severity prediction using a graph neural network framework." IEEE Access 12 (2024): 37540- 37556.

[2] Zohra, Ennaji Fatima, Knouzi Maryam, and Elalaoui Elabdallaoui Hasna. "Accident Severity Prediction using Machine Learning: A case study on the US Accidents Dataset." In 2023 17th International Confer- ence on Signal-Image Technology Internet-Based Systems (SITIS), pp. 242-246. IEEE, 2023.

[3] Shim, Junseok, Jaeseung Yu, and Kibeom Lee. "Integrated Risk Grid Map for Collision Avoidance and Mitigation Maneuvers of Autonomous Vehicle." IEEE Access (2025).

[4] Aziz, Kamran, Feng Chen, Inamullah Khan, Shabir Hussain Khahro, Abdul Malik Muhammad, Zubair Ahmed Memon, and Afaq Khattak. "Road traffic crash severity analysis: a Bayesian-optimized dynamic ensemble selection guided by instance

hardness and region of competence strategy." IEEE Access (2024).

[5] Hu, Lin, Haibo Li, Ping Yi, Jing Huang, Miao Lin, and Hong Wang. "Investigation on AEB key parameters for improving car to two-wheeler collision safety using in-depth traffic accident data." IEEE Transactions on Vehicular Technology 72, no. 1 (2022): 113-124.

[6] Zhang, Chen, Linjia Li, Aihui Pei, Tao Li, Junli Yuan, Xiao Feng, and Yuxi Zheng. "Severity Analysis of Factors Contributing to Two-vehicle Crashes Based on Interpretability Theory." IEEE Access (2024).

[7] Hu, Yaqi, Songmiao Zheng, Zirui Zhang, Siming Wang, Dongdong Ye, Maoqiang Wu, Xiaohuan Li, and Rong Yu. "Leveraging LLMs in Cloud-Edge Networks for Traffic Risk Prediction and Accident Severity Analysis." IEEE Transactions on Network Science and Engineering (2025).

[8] Manzoor, Mubariz, Muhammad Umer, Saima Sadiq, Abid Ishaq, Saleem Ullah, Hamza Ahmad Madni, and Carmen Bisogni. "RFCNN: Traffic accident severity prediction based on decision level fusion of machine and deep learning model." IEEE Access 9 (2021): 128359-128371.

[9] Bosia, Leonardo, Luca Manganotto, Marco Anghileri, Stefano Dolci, Paolo Astori, and Cing-Dao Kan. "Real-Time Collision Mitigation Strategies for Autonomous Vehicles." IEEE Transactions on Intelligent Transportation Systems 25, no. 10 (2024): 13483-13493.

[10] Kashifi, Mohammad Tamim. "Deep Hybrid Attention Framework for Road Crash Emergency Response Management." IEEE Transactions on Intelligent Transportation Systems 25, no. 8 (2024): 8807-8818. [11]Oad, Rahul, Ali Irtaza Sayani, and Shadi Banitaan. "Predicting Severity of US Traffic Accidents: A Machine Learning Approach." In 2024 IEEE International Conference on Electro Information Technology (eIT), pp. 679-685. IEEE, 2024.

[12] Aboulola, Omar Ibrahim, Ebtisam Abdullah Alabdulqader, Aisha Ahmed AlArfaj, Shtwai Alsubai, and Tai-Hoon Kim. "An automated approach for predicting road traffic accident severity using transformer learning and explainable AI technique." IEEE Access 12 (2024): 61062-61072.

[13] Ragam, Prashanth, Namana Sri Rahul, Pandi Balaji, Sankati Jyothi, and Pavithra Danaveni. "An Intelligent Machine Learning Approach for Prediction of Road Accident Severity." In 2025 3rd IEEE International Conference on Industrial Electronics: Developments Applications (ICIDeA), pp. 1-6. IEEE, 2025.

[14] Kumar, V. Aswin, GR Habi Radhan, R. S. Amshavalli, R. Rajalakshmi, and N. Jeenath Shafana. "Machine Learning-based Predictive Frame- work for Road Accident Severity Classification and Real- Time Risk Assessment." In 2025 International Conference on Visual Analytics and Data Visualization (ICVADV), pp. 1038-1044. IEEE, 2025.

[15] Paul, Joy, Zerin Jahan, Kazi Fahim Lateef, Md Robiul Islam, and Sagor Chandro Bakchy. "Prediction of road accidents and severity of Bangladesh applying machine learning techniques." In 2020 IEEE 8th R10 Humanitarian Technology Conference (R10-HTC), pp. 1-6. IEEE, 2020.

[16] Shawon, Ashadullah, and Akramul Azim. "Advancing Road Safety: Road Accident Severity Prediction Using Deep Learning Models." In 2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC), pp. 3575-3580. IEEE, 2024.

[17] Ahmed, Shakil, Md Akbar Hossain, Md Mafijul Islam Bhuiyan, and Sayan Kumar Ray. "A comparative study of machine learning algorithms to predict road accident severity." In 2021 20th International Conference on Ubiquitous Computing and Communications (IUCC/CIT/DSCI/SmartCNS), pp. 390-397. IEEE, 2021.

[18] Alagarsamy, Saravanan, P. Nagaraj, B. Srikanth, Ch Vamsi Krishna, G. Bharath, and Sv Sai Kalyan. "A Novel Machine Learning Technique for Predicting Road Accidents." In 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), pp. 1547-1551. IEEE, 2023.

[19] Pei, Yulong, Yuhang Wen, and Sheng Pan. "Road traffic accident risk prediction and key factor identification framework based on explainable deep learning." IEEE Access (2024).

[20] Kaliraja, C., D. Chitradevi, and Anju Rajan. "Predictive analytics of road accidents using machine learning." In 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), pp. 1782-1786. IEEE, 2022.