Optimized feature selection approaches for accident classification to enhance road safety

Mummaneni Sobhana, Gnana Siva Sai Venkatesh Mendu, Nihitha Vemulapalli, Kushal Kumar Chintakayala

Department of Computer Science and Engineering, V. R. Siddhartha Engineering College, Vijayawada, India

Article Info

ABSTRACT

In the modern era, the issue of road accidents has become an increasingly critical global concern, requiring urgent attention and innovative solutions. This investigation has compiled an extensive dataset of 10,356 accident occurrences that occurred between the years 2018 and 2022 in Ernakulam district. By utilizing advanced feature selection methodologies, such as genetic algorithm and coyote optimization, this research has identified pivotal accident determinants. The study harnesses the potential of deep learning techniques, encompassing recurrent neural network (RNN), gated recurrent unit (GRU), long short-term memory (LSTM), and multilayer perceptron (MLP) for classifying accidents according to severity (categorized as fatal, grievous, and severe). Eight predictive models are trained using the dataset, and the top two are ensembled. Integrating deep learning and optimization strategies, this research aims to create a robust accident classification system. The system will help in developing proactive policies that can reduce the frequency and severity of accidents in Ernakulam district
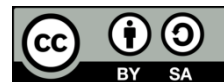
Corresponding Author:

Mummaneni Sobhana
Department of Computer Science and Engineering, V. R. Siddhartha Engineering College
Vijayawada 520007, India
Email: sobhana@vrsiddhartha.ac.in

1. INTRODUCTION

India's road safety crisis is evident with a nine-fold increase in accident fatalities from 1970 to 2013 [1]. Vulnerability is pronounced among economically active individuals aged 30-59, with males facing higher accident risks. Accidents peak during May-June and December-January, predominantly between 9 AM and 9 PM. Driver error persists as a significant factor, and regional disparities contribute to varying fatality rates among states. India's fatality rates surpass those of developed countries, underscoring the urgency for comprehensive road safety measures, effective policies, and political commitment.

Road accident classification is a pivotal field of study, focusing on the categorization of accidents according to their severity, root causes, and contributory elements [2]. This analysis holds profound importance for a multitude of purposes, notably in the development of robust road safety strategies, precise insurance assessments, and refined law enforcement approaches. Accidents are stratified into various classes, ranging from minor to fatal, contingent upon the nature of injuries and property damage sustained. The classification process considers factors such as weather conditions, driver behavior, road type, and vehicle speed. Statistical models, machine learning methods, and data analysis play a crucial role in revealing patterns and predicting trends in accidents. Precise accident classification empowers authorities to allocate

resources efficiently, enact targeted safety measures, evaluate risk, set insurance premiums, and ultimately foster safer road environments, culminating in the preservation of lives.

Choosing the right factors for road accident classification is vital [3]. Picking the most important variables from the many available like driver behavior, road type, and weather is crucial for building accurate predictive models. It boosts model performance, making it more efficient and easier to understand. Optimizing feature selection not only enhances classification accuracy but also aids in directing resources and efforts toward improving road safety. It's a significant step in reducing the social and economic impact of accidents.

The coyote optimization algorithm (COA) emerges as a novel metaheuristic approach, drawing inspiration from the behavioral patterns of coyotes (Canis latrans) [4]. COA, employing a population-based approach, organizes coyotes into packs representing potential solutions for optimization problems. Mimicking the cost function in coyotes' social conditions, COA fosters adaptation, interaction, and diversity, effectively navigating high-dimensional spaces for efficient feature selection. COA introduces mechanisms for birth, death, and cultural exchange led by an alpha coyote within each pack, showcasing superior performance in numerical evaluations and statistical tests compared to other nature-inspired metaheuristics.

In the domain of feature selection, genetic algorithms (GAs) represent a widely used optimization technique [5]. Drawing inspiration from natural selection and evolution, GAs mimic genetic variation, selection, and recombination to identify informative features within datasets. Starting with a population of potential feature subsets, GAs evolve them iteratively over multiple generations. Through selection, crossover, and mutation operations, GAs promote the survival of feature subsets demonstrating superior performance, ultimately leading to the discovery of optimal or near-optimal feature combinations. Their versatility and adaptability make GAs valuable for optimizing machine learning models, improving data analysis, and enhancing the efficiency of complex problem-solving tasks.

The paper is organized into five sections. The first section shows the introduction part. The second section is literature review, focused on existing research related to our study. The third section is method, which details our approach and techniques applied. The fourth section is results and discussion, which presents our findings. Finally, the conclusion section summarizes the findings and future work.

2. LITERATURE REVIEW

Sangare *et al.* [6] utilized a 2017 road traffic accident (RTA) dataset from data.govt.uk to develop a hybrid forecasting model. Combining gaussian mixture models (GMM) with support vector classifiers (SVC), the model significantly improves accuracy, showcasing potential for urban traffic accident forecasting. Wu *et al.* [7] employed crash prediction models-genetic algorithms (CPM-GAs) to forecast crashes based on road geometry and traffic data, integrating traditional CPM-GAs with machine learning models for enhanced accuracy. The findings underscore the superior accuracy of machine learning models in crash prediction. Devaraj *et al.* [8] analyzed road accidents in Kerala, determining severity levels based on contributory factors through year-wise, day-wise, and district-wise assessments. Utilizing a decision tree algorithm, accidents are classified into severe, medium, or low severity, aiding in identifying high-risk locations.

Labib *et al.* [9] addressed road accidents in Bangladesh, employing K-nearest neighbors (KNN), AdaBoost, naïve Bayes, and decision tree for severity classification. AdaBoost outperforms, achieving 80% accuracy, and the study explores factors like road class, junction type, surface condition, time, and vehicle type, offering insights to reduce accidents. Gilani *et al.* [10] investigated accident severity in Rasht City, utilizing logistic regression and artificial neural networks (ANN). The logistic regression model achieves 89.17% accuracy, identifying variables impacting severity such as time, weather, and vehicle type. The ANN model performs even better, with 98.9% accuracy, emphasizing the importance of vehicle quality and visibility measures to reduce accidents. Adanu *et al.* [11] studied injury severity in Alabama interstate crashes using random parameters multinomial logit modeling. Covering diverse factors, the study distinguishes variations in injury determinants between urban and rural, single-vehicle, and multi-vehicle crashes, providing insights for targeted road safety measures.

Islam *et al.* [12] examined serious traffic incidents involving individuals aged 15 to 44 in Al-Ahsa, Saudi Arabia, using descriptive analyses. Classification and regression tree (CART) and logistic regression models reveal this age group's involvement in severe crashes with higher injuries and fatalities, with CART highlighting specific scenarios. Santos *et al.* [13] analyzed accident data from Setúbal, Portugal (2016-2019) to identify factors influencing accident severity, utilizing machine learning algorithms like clustering, logistic regression, decision trees, and random forests, successfully identifying key factors for fatal and non-fatal accidents. Omari *et al.* [14] used fuzzy logic and geographic information system (GIS) to predict traffic accident hotspots in Irbid City, Jordan, analyzing data from 2013 to 2015. Identifying eight hotspots using

weighted overlay and fuzzy overlay methods with analytical hierarchy process (AHP), this research successfully reveals high-risk areas.

Bokaba *et al*. [15] examined machine learning techniques using real RTA data from Gauteng, South Africa. Evaluating classifiers, including logistic regression, naïve Bayes, KNN, random forest, AdaBoost, and support vector machine (SVM), the study finds the random forest classifier most effective with chained equations for multiple imputations. Santos *et al*. [16] introduced a novel approach to enhance road safety using a driver simulator model, analyzing factors like speed, curve radius, and time of day. The study provides valuable insights into road accidents, suggesting potential improvements in road safety through informed risk factor assessments. Ashraf *et al*. [17] investigated road accidents in South Korea, identifying factors such as traffic volume, limited road expansion, increasing passenger cars, safety violations, and driver characteristics through rainfall and accident data analysis.

Mesquitela *et al.* [18] tackled urban traffic accidents using smart city data and data fusion, combining accident data, weather conditions, and local reports through big data analytics. Geo-referenced accident hotspots are identified via kernel density and hot spot analysis (Getis-Ord Gi*) in ArcGIS Pro, aiding municipalities in understanding factors influencing accident severity. Sathiyaraj *et al.* [19] addressed traffic challenges in growing metropolitan areas, proposing the smart traffic prediction and congestion avoidance system (s-TPCA) with poisson distribution for vehicle arrival prediction. The system combines traffic recognition, forecasting, and congestion prevention, resulting in 20% higher fuel conservation compared to existing systems.

Olugbade *et al.* [20] highlighted road accidents, advocating artificial intelligence (AI) and machine learning for automatic incident detection systems, emphasizing route optimization and traffic management. The study offers insights into emerging trends and challenges, aiding road transport system planning and management. Satria *et al.* [21] aimed to enhance traffic safety, introducing the integrated nested laplace approximation-conditional autoregressive (INLA-CAR) model to assess crash severity on a highway section, emphasizing spatial correlation and average annual daily traffic influence. Sobhana *et al.* [22] introduced a framework for urban traffic crash analysis using deep learning techniques. It evaluates models like multilayer perceptron (MLP), recurrent neural network (RNN), long short-term memory (LSTM), and gated recurrent unit (GRU) to classify the severity of road accidents in Vijayawada, India. LSTM emerges as the most accurate model, achieving 93.07% accuracy.

3.   METHOD

The proposed methodology involves several key modules in the context of Ernakulam. Firstly, the data collection module acquires relevant accident data from 2018 to 2022. Next, the feature selection module applies advanced techniques like GA and coyote optimization. Then, the classification module employs deep learning algorithms, including RNN, GRU, LSTM, and MLP, to categorize severity of road accidents. Eight predictive models are trained and then best two models are ensembled. The ultimate objective is to develop a robust accident classification system for the implementation of proactive safety policies in Ernakulam district. Figure 1 shows the proposed methodology diagram.

To address the challenge of missing data, the study employs multiple imputation, a technique that generates diverse datasets with unique imputations. Each dataset undergoes independent analysis, and the outcomes are subsequently merged, culminating in a more robust result. This approach meticulously considers the variability associated with missing data, fostering a nuanced understanding of the dataset. Special attention is given to potential biases, particularly when dealing with missing categorical values. Statistical procedures are then applied to aggregate the results, enhancing the reliability of subsequent analyses and modeling tasks in complex datasets.

The study utilizes one-hot encoding to transform categorical data into a numerical format compatible with machine learning algorithms. Initially, categorical variables are identified, and the unique categories within each variable are determined. For each category, a binary column is created, where '1' signifies the presence of the category, and '0' indicates its absence. These binary columns are appended to the original dataset, effectively expanding the feature space. Subsequently, the original categorical columns are dropped, ensuring the dataset's readiness for machine learning applications.

3.1. Genetic algorithm based selection of features

The application of the GA for feature selection is a systematic method that aims to enhance model performance by identifying crucial features within a dataset. Beginning with the definition of input parameters, including chromosome length (L), number of features to be selected (N), mutation rate, and maximum generations (Gmax), the algorithm initializes a population of potential feature subsets. Through iterative processes, it evaluates and selects chromosomes based on their positive contributions to model fitness. Crossover operations blend genetic information of selected parents, generating offspring, while mutation introduces small variations. Least fit individuals are replaced with offspring in subsequent

generations. The algorithm concludes by extracting the chromosome with the highest fitness, representing an optimal feature subset for enhanced model accuracy and efficiency.

3.2. Coyote optimization based selection of features

A metaheuristic technique called coyote optimization is used in selecting features to improve the process of finding the best subset of features for classification problems. The algorithm employs a population-based approach, initializing parameters such as population size, feature space dimension, and mutation rates. Through iterative cycles of exploration and exploitation, the algorithm evaluates candidate feature subsets. During the exploration phase, features undergo mutation, while the exploitation phase involves local search operations. Communication among coyotes within the population facilitates information sharing on promising features. The algorithm incorporates termination conditions, such as reaching a specified maximum number of generations. The final result extraction selects the feature subset with the highest fitness, effectively optimizing feature selection for improved classification.

3.3. Ensembling

The process begins with the input of a preprocessed road accident dataset, and the ultimate goal is to predict the severity of road accidents. First, the dataset is divided into training data and testing data using 80:20 ratio. Feature selection is then carried out through a two-step process involving GA and coyote optimization. The relevant features are initially identified using GA, and the selection is further refined using COA to obtain the optimal feature subset. Subsequently, deep learning algorithms like GRU, RNN, LSTM, and MLP are trained on the dataset using the identified optimal feature subset. Testing data is used to evaluate the accuracy of every model. To enhance performance, the top two models based on accuracy are selected, and an ensemble model is created by stacking these chosen models. Finally, predictions on accident severity are made using the ensemble model. The study evaluates the ensemble model's performance by employing accuracy, model loss and model accuracy curves. The accuracy of the model is evaluated using (1):

$$Accuracy = \frac{True\ Positives + True\ Negatives}{Total\ Samples} \qquad (1)$$
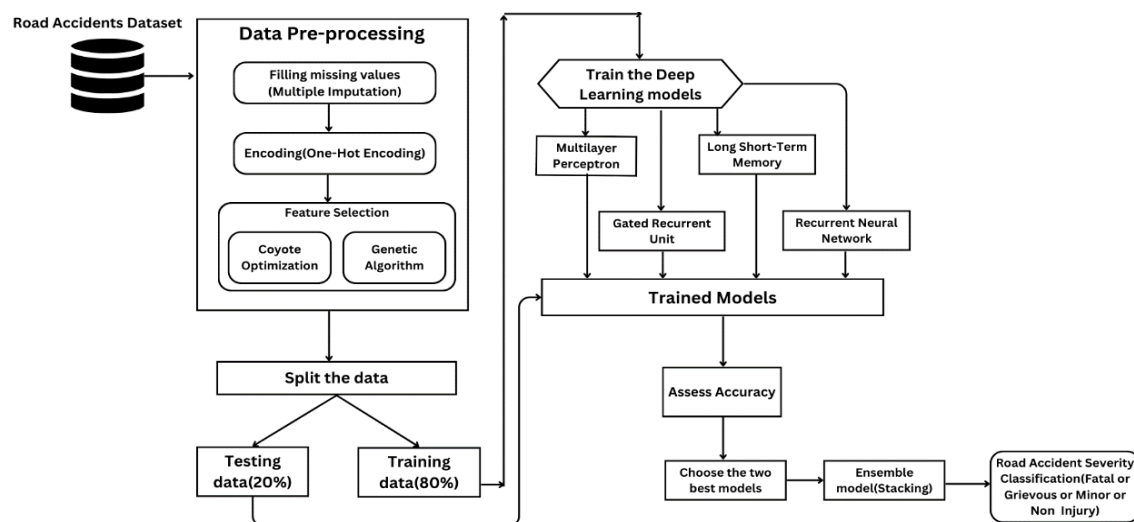

Figure 1. Methodology diagram

4. RESULTS AND DISCUSSION

This study investigates the impact of GA and COA in feature selection for an ensemble classification model predicting road accident severity. While previous research explores various feature selection techniques, there is a gap in understanding the collective influence of GA and COA on ensemble models. Existing studies predominantly focus on individual methods, overlooking the synergistic effects of these two algorithms. This research addresses this gap by examining how GA and COA enhance the predictive performance of an ensemble model for road accident severity classification. The feature selection process conducted by GA and COA on the original 46-feature dataset revealed insightful outcomes. GA identified 15 features, whereas COA refined the set to 14 features. Table 1 illustrates the detailed overview of the features chosen by each algorithm.

Table 1. Features selected before and after optimization

Initial set of attributes (46)			Features selected using GA (15)	Features selected using COA (14)
Zone	Minor	Divider	Sections	Range
Range	Driver	Spot accident	Accident type	PS name
District	Passenger	Speed limit	Death	Firno
Subdivision	Pedestrian	Weather	Grievous	Time accident
Circle	Cyclist	Road no	Longitude	Sections
PS name	Other persons	Road surface	Divider	Death
Firno	Motorised	T - junction	Weather	Grievous
Date report	Non motorised	Road chainage	Road surface	Driver
Date accident	Latitude	Hit run	T – junction	Pedestrian
Time report	Longitude	Collision	Road chainage	Cyclist
Time accident	Place of occurance	Type road	Hit run	Non motorised
Sections	Type area	Cause accident	Cause accident	Latitude
Accident type	On going road works	Road features	Road features	Longitude
Death	City/town/village	Visibility	Visibility	Place of occurance
Grievous	Lanes road		Traffic control	
Traffic control				
Vehicle type				

Table 2 displays the accuracies of both deep learning models and hybrid models achieved through feature selection using GA and coyote optimization. Among these models, the two highest-performing ones, COA+RNN and COA+LSTM, stand out with an impressive accuracy of 99.77%. These top-performing models have been chosen for ensemble, contributing to a comprehensive understanding of the enhanced performance achieved through feature optimization.

Table 2. Accuracy of the models without optimization and with optimization

Models	Without optimization	GA	Coyote optimization
MLP	83.84	99.77	99.74
RNN	72.15	96.58	99.77
LSTM	72.12	96.84	99.77
GRU	72.15	96.68	99.73

4.1. Ensemble model

The ensemble model results from stacking COA+RNN and COA+LSTM, combining their predictive capabilities. This approach leverages the strengths of each model to enhance overall performance, achieving a robust and highly accurate prediction with improved results. Figure 2 shows the model loss and accuracy curves of the ensembled model.
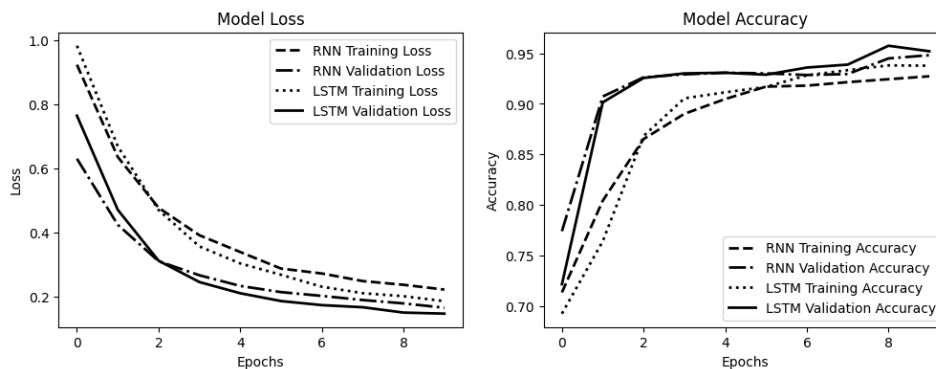

Figure 2. Loss and accuracy curves of ensemble model

In classification tasks, a confusion matrix serves as a valuable instrument, encapsulating the performance of a model. It offers a thorough breakdown of true positives, true negatives, false positives, and false negatives, facilitating the evaluation of F1-score, precision, accuracy, and recall. Figure 3 illustrates the confusion matrix for the ensemble model.
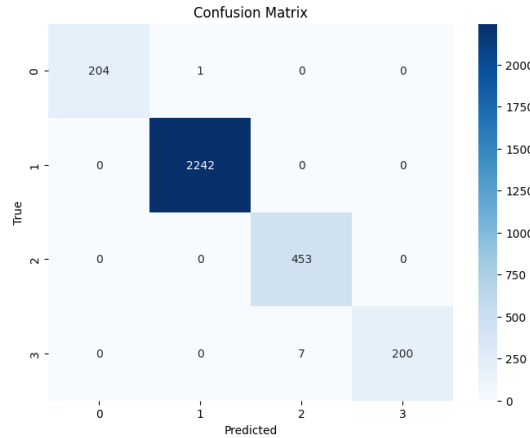
Figure 3. Confusion matrix of ensemble model

4.2. Comparision with other models

Our findings suggest a distinctive strength in the ensemble model (COA-RNN+COA-LSTM) for road accident severity classification, showcasing an exceptional accuracy of 99.77%. This surpasses the accuracy reported in various alternative models, such as RNN, random forest and convolutional neural network (RFCNN), SVM, KNN, MLP, logistic regression, simple CART, and partial decision trees classifier (PART) as illustrated in Table 3. Notably, our ensemble model outperforms these models, indicating its efficacy in accurately predicting accident severity.

Table 3. Comparision with other models

S. No	Author	Dataset	Methodology	Metric score (Accuracy)
1	Sameen and Pradhan [23]	North-South expressway (NSE)	RNN	71.77%
2	Manzoor *et al.* [24]	Car accident dataset (USA)	RFCNN	RFCNN=99.1%
3	Vaiyapuri and Gupta [25]	Web data from data.gov.in	SVM, RF, KNN, MLP, Logistic regression	SVM=60% RF=88% KNN=85% MLP=90% Logistic regression=54%
4	Wahab and Jiang [26]	Road traffic crash files	Simple CART, PART and MLP	Simple CART=73.81% PART=73.45% MLP=72.16%

4.3. Accident hotspots

Ernakulam district is under consideration, encompassing over 10,000 accident records spanning from 2018 to 2022. All these accident locations are mapped on the Ernakulam district map using ArcGIS. Areas with a higher number of accidents are highlighted in intense red, while those with fewer accidents are marked with a lighter red color. Figure 4 illustrates the Ernakulam District Map, showcasing the representation of accident hotspots.
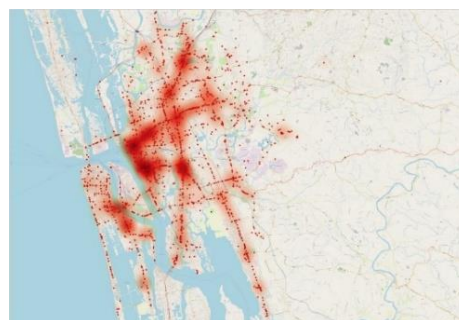
Figure 4. Accident hotspots in Ernakulam district

5. CONCLUSION

While our study comprehensively explored feature selection using GA and COA and demonstrated the robustness of the ensemble model, certain limitations should be acknowledged. The assumption that the identified features accurately represent the complexities of road accidents may be influenced by variations in data collection methods or regional characteristics, impacting generalizability. In Ernakulam district, our focused exploration of road accident challenges led to the development of an ensemble model (COA-RNN+COA-LSTM) achieving 99.77% accuracy in categorizing accident severity. While our study is confined to Ernakulam, it provides valuable insights into local road safety. Future efforts will concentrate on refining models through algorithm fine-tuning, optimization, and dataset expansion. Our plans involve extending the system beyond Ernakulam to contribute to a substantial reduction in accidents on a broader scale, alleviating financial burdens for victims and their families.

REFERENCES

[1] S. K. Singh, "Road traffic accidents in India: issues and challenges," *Transportation Research Procedia*, vol. 25, pp. 4708–4719, 2017, doi: 10.1016/j.trpro.2017.05.484.

[2] B. Kumeda, F. Zhang, F. Zhou, S. Hussain, A. Almasri, and M. Assefa, "Classification of road traffic accident data using machine learning algorithms," in *2019 IEEE 11th International Conference on Communication Software and Networks, ICCSN 2019*, 2019, pp. 682–687, doi: 10.1109/ICCSN.2019.8905362.

[3] Z. O. Hamad, "Review of feature selection methods using optimization algorithm," *Polytechnic Journal*, vol. 12, no. 2, pp. 203–214, 2023, doi: 10.25156/ptj.v12n2y2022.pp203-214.

[4] S. S. S. Reddy, A. Kumar, K. Z. Ghafoor, V. P. Bhardwaj, and S. Manoharan, "CoySvM-(GeD): coyote optimization-based support vector machine classifier for cancer classification using gene expression data," *Journal of Sensors*, vol. 2022, pp. 1–9, 2022, doi: 10.1155/2022/6716937.

[5] J. Zhou and Z. Hua, "A correlation guided genetic algorithm and its application to feature selection," *Applied Soft Computing*, vol. 123, 2022, doi: 10.1016/j.asoc.2022.108964.

[6] M. Sangare, S. Gupta, S. Bouzefrane, S. Banerjee, and P. Muhlethaler, "Exploring the forecasting approach for road accidents: Analytical measures with hybrid machine learning," *Expert Systems with Applications*, vol. 167, pp. 1–28, 2021, doi: 10.1016/j.eswa.2020.113855.

[7] P. Wu, X. Meng, and L. Song, "A novel ensemble learning method for crash prediction using road geometric alignments and traffic data," *Journal of Transportation Safety and Security*, vol. 12, no. 9, pp. 1128–1146, 2020, doi: 10.1080/19439962.2019.1579288.

[8] D. Devaraj, D. Chandrasekaran, B. Pandian, B. Binitha, M. Dipikhasre, and R. Anil, "Road accident analysis in Kerala and location-based severity level classification using decision tree algorithm," *Paideuma Journal*, vol. 14, pp. 41–50, 2021.

[9] M. F. Labib, A. S. Rifat, M. M. Hossain, A. K. Das, and F. Nawrine, "Road accident analysis and prediction of accident severity by using machine learning in Bangladesh," in *2019 7th International Conference on Smart Computing & Communications (ICSCC)*, 2019, pp. 1–5, doi: 10.1109/ICSCC.2019.8843640.

[10] V. N. M. Gilani, S. M. Hosseinian, M. Ghasedi, and M. Nikookar, "Data-driven urban traffic accident analysis and prediction using logit and machine learning-based pattern recognition models," *Mathematical Problems in Engineering*, vol. 2021, pp. 1–11, 2021, doi: 10.1155/2021/9974219.

[11] E. K. Adanu, W. Agyemang, R. Islam, and S. Jones, "A comprehensive analysis of factors that influence interstate highway crash severity in Alabama," *Journal of Transportation Safety and Security*, vol. 14, no. 9, pp. 1552–1576, 2022, doi: 10.1080/19439962.2021.1949414.

[12] M. K. Islam, U. Gazder, R. Akter, and M. Arifuzzaman, "Involvement of road users from the productive age group in traffic crashes in Saudi Arabia: an investigative study using statistical and machine learning techniques," *Applied Sciences*, vol. 12, no. 13, pp. 1–22, 2022, doi: 10.3390/app12136368.

[13] D. Santos, J. Saias, P. Quaresma, and V. B. Nogueira, "Machine learning approaches to traffic accident analysis and hotspot prediction," *Computers*, vol. 10, no. 12, pp. 1–15, 2021, doi: 10.3390/computers10120157.

[14] A. A. -Omari, N. Shatnawi, T. Khedaywi, and T. Miqdady, "Prediction of traffic accidents hot spots using fuzzy logic and GIS," *Applied Geomatics*, vol. 12, no. 2, pp. 149–161, 2020, doi: 10.1007/s12518-019-00290-7.

[15] T. Bokaba, W. Doorsamy, and B. S. Paul, "Comparative study of machine learning classifiers for modelling road traffic accidents," *Applied Sciences*, vol. 12, no. 2, pp. 1–18, 2022, doi: 10.3390/app12020828.

[16] M. I. Santos, P. T. M. S. Oliveira, and A. P. C. Larocca, "Investigation the influence of risk factors on the occurrence of road accidents using the driver performance model," *Transportation in Developing Economies*, vol. 8, no. 1, pp. 1–11, 2022, doi: 10.1007/s40890-021-00148-x.

[17] I. Ashraf, S. Hur, M. Shafiq, and Y. Park, "Catastrophic factors involved in road accidents: underlying causes and descriptive analysis," *PLoS ONE*, vol. 14, no. 10, pp. 1–29, 2019, doi: 10.1371/journal.pone.0223473.

[18] J. Mesquitela, L. B. Elvas, J. C. Ferreira, and L. Nunes, "Data analytics process over road accidents data—a case study of Lisbon City," *ISPRS International Journal of Geo-Information*, vol. 11, no. 2, pp. 1–18, 2022, doi: 10.3390/ijgi11020143.

[19] R. Sathiyaraj, A. Bharathi, S. Khan, T. Kiren, I. U. Khan, and M. Fayaz, "A genetic predictive model approach for smart traffic prediction and congestion avoidance for urban transportation," *Wireless Communications and Mobile Computing*, vol. 2022, pp. 1–12, 2022, doi: 10.1155/2022/5938411.

[20] S. Olugbade, S. Ojo, A. L. Imoize, J. Isabona, and M. O. Alaba, "A review of artificial intelligence and machine learning for incident detectors in road transport systems," *Mathematical and Computational Applications*, vol. 27, no. 5, pp. 1–24, 2022, doi: 10.3390/mca27050077.

[21] R. Satria, J. A. -Valverde, and M. Castro, "Spatial analysis of road crash frequency using bayesian models with integrated nested laplace approximation (INLA)," *Journal of Transportation Safety and Security*, vol. 13, no. 11, pp. 1240–1262, 2021, doi: 10.1080/19439962.2020.1726542.

[22] M. Sobhana, N. Vemulapalli, G. S. S. V. Mendu, N. D. Ginjupalli, P. Dodda, and R. B. V. Subramanyam, "Urban traffic crash analysis using deep learning techniques," *Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Srodowiska*, vol. 13, no. 3, pp. 56–63, 2023, doi: 10.35784/iapgos.5350.

[23] M. I. Sameen and B. Pradhan, "Severity prediction of traffic accidents with recurrent neural networks," *Applied Sciences*, vol. 7, no. 6, pp. 1–17, 2017, doi: 10.3390/app7060476.

[24] M. Manzoor *et al.*, "RFCNN: traffic accident severity prediction based on decision level fusion of machine and deep learning model," *IEEE Access*, vol. 9, pp. 128359–128371, 2021, doi: 10.1109/ACCESS.2021.3112546.

[25] T. Vaiyapuri and M. Gupta, "Traffic accident severity prediction and cognitive analysis using deep learning," *Soft Computing*, pp. 1–13, 2021, doi: 10.1007/s00500-021-06515-5.

[26] L. Wahab and H. Jiang, "Severity prediction of motorcycle crashes with machine learning methods," *International Journal of Crashworthiness*, vol. 25, no. 5, pp. 485–492, 2020, doi: 10.1080/13588265.2019.1616885.

BIOGRAPHIES OF AUTHORS

Dr. Mummaneni Sobhana is currently working as an associate professor in the Department of Computer Science and Engineering, V. R. Siddhartha Engineering College, Vijayawada, India. She received Ph.D. degree in Computer Science and Engineering in 2018 from Krishna University. She has 16 years of teaching experience. Her research interests lie in areas such as artificial intelligence, machine learning, data analytics, cyber security, and software engineering. She published 35 papers in national and international journals, and published 7 patents. She can be contacted at email: sobhana@vrsiddhartha.ac.in.

Gnana Siva Sai Venkatesh Mendu is a final-year B.Tech. student, specializing in Computer Science and Engineering at V. R. Siddhartha Engineering College, Vijayawada, India. He is passionate about artificial intelligence and machine learning. He achieved Bronze in the Cisco NetAcad Riders 2023 competition. He holds the position of Institute of Electrical and Electronics Engineers (IEEE) chair of the geoscience and remote sensing society (GRSS) student chapter. He can be contacted at email: sivasaivenkatesh.m@gmail.com.

Nihitha Vemulapalli is a final-year B.Tech. student specializing in Computer Science and Engineering at V. R. Siddhartha Engineering College, Vijayawada, India. She is passionate about deep learning and has been recognized as a Google Women Engineers Scholar. Additionally, she holds the position of IEEE chair at the women in engineering (WIE) student chapter. She can be contacted at email: nihithavemulapalli@gmail.com.

Kushal Kumar Chintakayala is a final-year B.Tech. student, specializing in Computer Science and Engineering at V. R. Siddhartha Engineering College, Vijayawada, India. He is passionate about artificial intelligence and machine learning. He is IEEE core member of Computer Society of India (CSI) student chapter. He can be contacted at email: kushalkumarchintakayala12@gmail.com.