Advanced International Journal for Research (AIJFR)
E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

Review On Development of AI Based Road Accidents Prediction Through Traffic Image Analysis

Ms. Shital Narwade[1] , Mr. Rahul Bhandekar[2]

[1]PG Scholar , Computer Science & Engineering, Wainganga College of Engineering & Management, Nagpur, India
[2]Assistant Professor , Computer Science & Engineering, Wainganga College of Engineering & Management, Nagpur, India

Abstract
Traffic accidents are one of the primary reasons for death and serious injury. They are a great threat not only to the health of people but also to their lives. Even if these incidents happen due to numerous causes, some of which are internal to the driver and others external, they will still occur. When the visibility is poor during bad weather like rain, clouds, and fog, driving may become very hard and even hazardous. This project intends to provide an overview of advanced methods for traffic accident prediction by using machine learning algorithms and clustering techniques. The increasing number of vehicle collisions globally has got an impact on many aspects of human life. The aspects like causation study, traffic flow and the interaction between different factors have been mostly overlooked, though being significant. Also, the current traffic accident data is mainly used for data mining and basic statistical analysis, which results in a lack of understanding of the statistics and the trends. By classifying the road accident data, this project aims at minimizing the severity of further accidents through spotting the main contributing factors and formulating preventive measures. Machine learning algorithms are employed to analyze the data, detect the underlying patterns, predict the occurrence's intensity and disseminate the information quickly.

Keywords: Road accident data, Machine learning, K-means Clustering, Analysis, Visualization, prediction etc.

1. Introduction

The World Health Organization recently published a report with fatality statistics, which indicates an extremely high number of accidents every year on the roads around the globe. Unfortunately, road incidents can happen at any place. The rate at which vehicular traffic is increasing today is enormous, thus, leading to way more accidents on the roads. As a result, accident prediction has become one of the most important areas of research in transportation safety. The probability of traffic accidents is affected, to a great extent, by factors such as the shape of the road, the flow of traffic, the characteristics of the drivers, and the environment around. A whole lot of different types of research have been performed, including finding the dangerous spots or areas of concern, studying the nature and causes of road

accidents, and calculating the future accident occurrence, etc. The accidents' patterns are being investigated in some studies. Besides, road and weather conditions have been the main concerns too. There is no specific technique being applied by the traffic police for the detection of the locations having a high risk of accidents. Predicting road accidents is a necessity for organized traffic flow and management since lots of uncertainties exist in the causes of accidents like the weather, people, vehicles, and roads. Machine learning algorithms can deal with large-scaled categorization variables which could then be used for the detection of interesting patterns. It is scalable and can handle colossal data processing. The clustering method also plays a role in the analysis and visualization of data concerning vehicular accidents.

Mullard and Lass' system model for the human-induced environmental changes and instability systems categorically says that environmental, human, and vehicle defects are oftentimes the main reasons for traffic accidents [4]. The built and natural surroundings, as well as transportation systems, are all counted as environmental elements. Weather, smoky air, strong winds, and bad lighting are the most common environmental elements that have direct influence on road traffic accidents. The infrastructural factors, which fall under the environmental category, consist of road type, highway condition, highway lane type, poor road surface, inadequate road markings, inadequate traffic signs, and road layout. Among human factors are gender, age, education, human behaviors, driving skill, driving style, risk-taking and risky driving (which encompasses using alcohol and illegal drugs, speeding, jumping red lights, among others). The vehicle factors encompass its design, age, volume, and quality, as well as its technical requirements, safety features, and shortcomings in design.

2. Problem Identification

The World Health Organization has announced death statistics, which point to an extremely high number of accidents every year on roads all over the globe. Road accidents can take place anywhere. The number of vehicles on the road is growing very fast today, which is one of the main causes of more accidents. Consequently, predicting road traffic accidents is one of the most important areas of research in transportation safety. The likelihood of an accident occurring is largely determined by the road condition, the number of vehicles, the drivers, and their environment. Numerous studies have been conducted, among them the mapping of accident prone areas or “hot spots,” revealing the circumstances of road accidents, and forecasting the number of accidents. The reasons behind accidents are examined in some studies. The weather and the driver's visibility of the road are additional considerations. The police on the roads have no uniform way to spot a dangerous area. Predicting road crash incidents is vital for the organized and smooth running of traffic, as the major causes of crashes are so unpredictable—all including humans, vehicles, roads, weather, and other nonlinear variables. Computer learning. To reveal important patterns, algorithms could sort through an enormous number of classification parameters. It is not only scalable but also capable of handling large quantities of data. Furthermore, the clustering technique facilitates the assessment and visualization of the data concerning road accidents.

3. LITERATURE SURVEY

At present, road traffic accidents are a global concern that needs to be addressed by investigators, NGOs, car manufacturers, authorities, and the business community not just as a public health issue but also as a development problem.

Zhang et al., 2025, studies on deep-learning-based traffic accident anticipation using images and videos. They group methods into four families: image/video feature-based, spatiotemporal modeling, scene understanding, and multimodal fusion. The paper compares CNNs, RNNs, transformers, and hybrid models across benchmark datasets and synthetic scenarios. It highlights problems such as data scarcity, annotation cost, domain shift, imbalanced accident vs non-accident samples, and real-time deployment constraints. The authors emphasize the need for unified evaluation protocols, larger diverse datasets, explainable models, and integration with intelligent transportation systems and smart-city infrastructures.

Li et al., 2024, systematically survey machine-learning approaches used for traffic accident analysis, severity prediction, and hotspot identification. The review covers classical statistical models, tree-based ensembles, support vector machines, neural networks, and deep-learning architectures. It compiles applications from freeway crashes, urban intersections, and regional networks, and compares performance under different feature sets such as traffic flow, weather, geometric design, and driver information. The authors stress recurring issues like class imbalance, limited spatial–temporal generalization, and lack of transferability between regions. They also underline the importance of interpretable models, integration with real-time traffic management platforms, and combining heterogeneous data sources such as sensors, crowdsourcing, and connected vehicles.

Yang et al., 2025, conduct a systematic literature review of 78 papers using machine learning and deep learning for road accident prediction. They categorize models into regression, tree-based ensembles, SVMs, probabilistic models, neural networks, and hybrid deep-learning architectures. The paper discusses feature engineering from traffic volume, road geometry, weather, and human factors, and summarizes performance metrics used across studies. The review finds that deep-learning methods can capture complex nonlinear relations but often require large, high-quality datasets that many regions lack. It also identifies gaps in handling uncertainty, explainability, and real-time deployment. The authors propose a research roadmap emphasizing interpretable deep models, uncertainty quantification, and integration with policy and infrastructure planning.

Shaik et al., 2021, focus specifically on neural-network-based methods for predicting road traffic accident severity. The review compares shallow neural networks, deep feedforward architectures, and hybrid models combining NNs with fuzzy logic or optimization algorithms. It summarizes datasets used across countries, typical input features (road conditions, vehicle factors, environmental variables, and driver attributes), and evaluation criteria such as accuracy, F1-score, and AUC. The authors conclude that deep networks can outperform traditional models when sufficient data are available, but they are sensitive to imbalance and require careful hyperparameter tuning. They highlight research needs in explainable neural networks, transfer learning across regions, and incorporating temporal sequences and spatial correlations into severity prediction.

Adewopo et al., 2023, present an intensive review of computer-vision-based action recognition techniques applied to traffic accident detection in smart cities. They survey deep-learning models using surveillance cameras, dashcams, drones, and on-board vehicle cameras. The paper discusses two-stream CNNs, 3D CNNs, CNN–LSTM hybrids, transformer-based architectures, and semi-Markov models for spatiotemporal pattern recognition in video. It also analyzes benchmark datasets, sensor modalities, and evaluation metrics. The authors identify challenges such as limited annotated accident footage, complex multi-class event localization, high computational cost, and biases in existing datasets. They propose future research directions involving better benchmark creation, multimodal fusion, transfer learning, and lightweight architectures suitable for real-time deployment in urban traffic-monitoring systems.

Fang et al., 2024, present the first dedicated survey on vision-based traffic accident detection (Vision-TAD) and traffic accident anticipation (Vision-TAA) in the deep-learning era. They summarize detection methods that localize accidents after they occur and anticipation methods that predict accidents before impact from video streams. The survey organizes approaches by model family (CNNs, RNNs, 3D CNNs, graph networks, transformers) and by task (frame-level classification, segment-level localization, risk prediction). It reviews public datasets, evaluation protocols, and real-world deployments. Key challenges highlighted include long-tail accident events, occlusions, complex interactions among vehicles, and the gap between controlled benchmarks and urban environments. The survey encourages research on interpretable models, uncertainty-aware risk scoring, and integration with advanced driver-assistance systems.

Chitraranjan et al., 2025, systematically review ego-centric, vision-based collision warning systems using deep learning. They analyze 31 studies proposing camera-based models to predict impending collisions from the vehicle's viewpoint. The review summarizes network architectures, including single-image risk prediction, temporal sequence models, and depth/optical-flow–aware designs. It evaluates datasets, labeling strategies for "risky" vs "safe" scenes, and validation practices. A key contribution is a risk-of-bias assessment using PROBAST, revealing common issues in study design, data leakage, and limited external validation. The authors conclude that while deep learning shows strong potential for collision warning, future work must address robustness to weather, lighting, sensor noise, and domain transfer from experimental datasets to everyday driving conditions.

Manguri et al., 2023, review computer-vision methods used in traffic monitoring, focusing on four main tasks: traffic density estimation, traffic sign detection and recognition, accident detection, and emergency vehicle detection. For accident detection, the paper summarizes techniques using object detection, motion analysis, and deep spatiotemporal models on CCTV feeds. It compares popular algorithms (YOLO, Faster R-CNN, SSD, optical-flow-based methods, and CNN–LSTM hybrids) and discusses their strengths and limitations under varying illumination, occlusions, and camera angles. The review also highlights dataset constraints, annotation challenges, and the need for more comprehensive benchmarks. The authors recommend integrating vision-based methods with IoT sensors and edge computing to achieve scalable, real-time intelligent transportation systems.

Silva et al., 2020, review the application of machine-learning techniques to road safety modeling, including accident frequency, severity, and risk-factor analysis. They compare traditional statistical

models (Poisson, negative binomial, logistic regression) with modern machine-learning approaches such as random forests, gradient boosting, SVMs, and neural networks. The review examines how different methods handle nonlinearities, interactions, and heterogeneous data, and summarizes reported gains in predictive performance. It also discusses interpretability tools, including variable importance and partial dependence plots, to explain model behavior. The authors stress the importance of data quality, spatial–temporal structure, and careful validation to avoid overfitting. They call for more applications combining ML with exposure, infrastructure planning, and policy evaluation.

Khan et al., 2025, presents an AI-centered review and experimental study on modeling road traffic accident severity using machine-learning techniques. The paper surveys prior works applying algorithms such as k-nearest neighbors, decision trees, AdaBoost, and neural networks to accident prediction and severity classification. It highlights advantages of ML in capturing complex dependencies between environmental, vehicular, and human-related factors compared with traditional models. Using a real accident dataset, the authors compare several ML classifiers for severity prediction and discuss feature importance for key risk factors. They emphasize the potential of AI to support policymakers and practitioners, while warning that imbalanced data, limited feature coverage, and lack of external validation still restrict real-world deployment.

Data mining techniques are important for assessing and projecting the value of traffic accident data in the future and for identifying patterns in the event elements that impact different metrics, according to multiple studies. Additionally, the enormous potential of information mining prediction approaches contributes significantly to avoiding and tracking the issues with road accident safety.

4. METHODOLOGY

During the training and testing phase, the system uses machine learning to determine the proposed model.

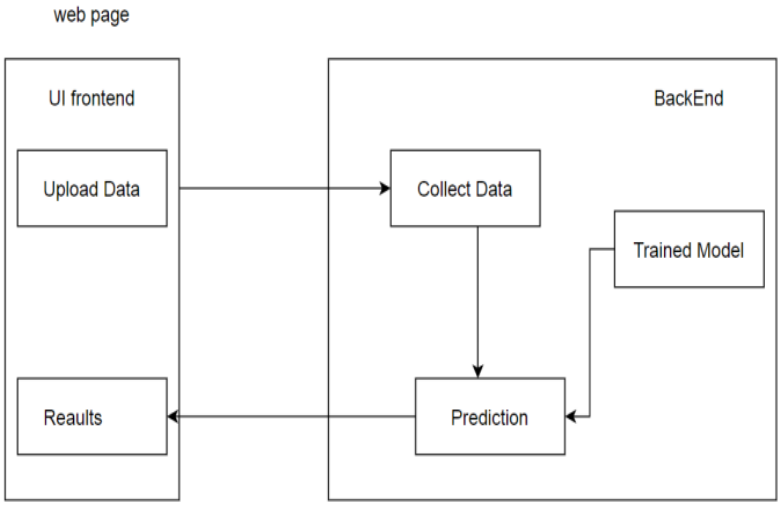

Fig.1. Architecture of system

• Data is the backbone of any data analysis and the most vital factor. The right type of data should be collected. The data content and the data structure need to be analyzed and understood thoroughly. The data for the analysis is taken from Kaggle and government sites.
• After data gathering, the next step is to analyze it. To analyze it, we need a tool which makes the work easy and efficient. We already had in mind to use Python for the coding part.
• The two libraries which contributed the most to the analysis are pandas and numpy. The main function of pandas is to perform data manipulation and analysis. In particular, it provides data structures and operations for manipulating numerical tables and time series. It provides high-performing, easily usable data structures and analysis tools.
• Numpy is the acronym for Numerical Python or Numeric python. It is a free software module that allows fast computation on arrays and matrices. Numpy is the core package for scientific computing with Python.
• Next let's talk about the algorithm we used. There are plenty of algorithms available to assist us with data analysis. The combination of machine learning and data analytics techniques is a win-win situation in this domain. The algorithm we chose is Regression Analysis.
• Logistic Regression Analysis is a collection of statistical methods for estimating the relationships among variables. It encompasses various techniques to model and analyze multiple variables, when the dependent variable is the focus and its affiliation with one or more independent variables is explored.

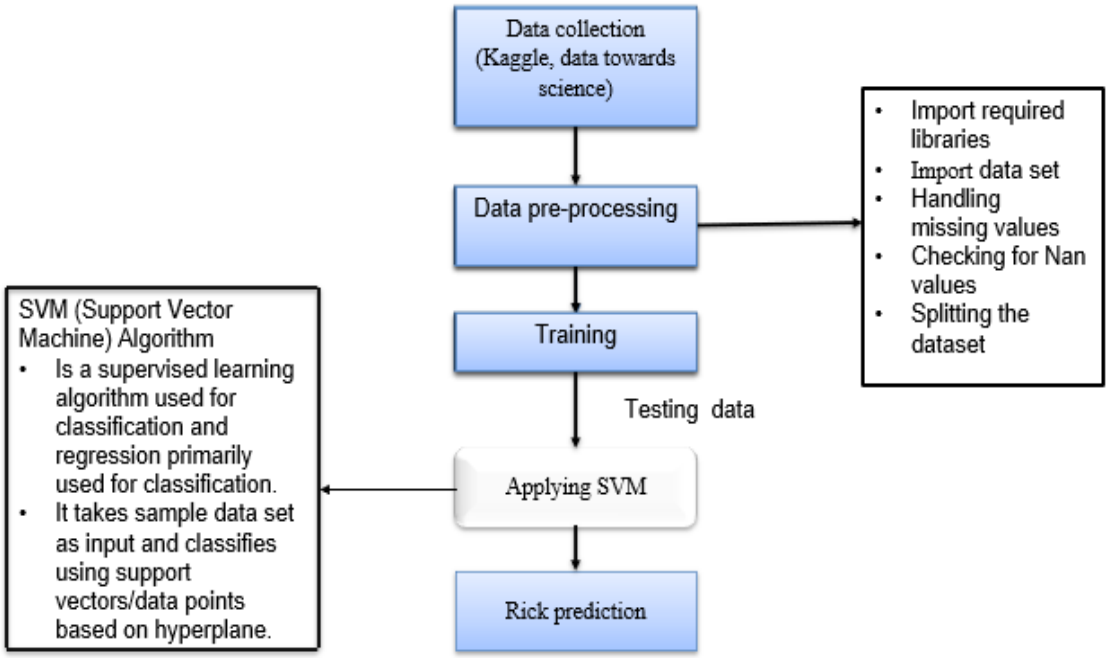

Fig.2. Flow diagram for Rule Mining

1. Data Collection

The first step is to obtain datasets related to accidents from various online open-source platforms like:

- Kaggle
- DataTowardScience
- Government open data portals

Datasets contain details about the weather, the condition of the road, the vehicles involved, the amount of traffic, type and severity of the accident, etc.

2. Data Pre-processing

- The raw data is being cleaned and formatted using various methods to make it ready for model training.

In the right-side box, the pre-processing steps are explained:

Steps Involved:

- Import the libraries needed (Pandas, NumPy, Scikit-learn, etc.)
- Load the dataset into the working space
- Deal with missing values (fill, remove, or interpolate missing entries)
- Look for NaN values to check data consistency
- Partition the dataset into training set and testing set
- This way, it is guaranteed that the input data is clean, free of errors, and appropriate for machine learning.

3. Training Phase

Once preprocessing is done, the refined dataset is employed to teach the machine learning model.

This phase facilitates the algorithm to discover the patterns, relationships, and dependencies in the data.

4. Utilizing the SVM (Support Vector Machine) Algorithm

- The SVM section clarifies the ML model that is applied:
- SVM is a supervised learning technique primarily for classifying the data, yet it can also do regression.
- The algorithm identifies the best hyperplane that separates the data into different classes and thus classifies the data.
- The decision boundary is formed based on support vectors (the critical data points that are closest to the boundary).
- SVM is applied to the testing dataset in this case to assess the model's capability and to categorize accident risk levels.

5. Risk Prediction

- This is the conclusion of the output stage.

The system predicts the following after SVM has been applied to the test data:

- The likelihood of an accident
- The seriousness of accidents that might happen
- Areas or conditions considered to be at high risk
- Thus, traffic authorities will be able to spot accident-prone areas and implement measures to prevent accidents.

5. TOOLS / PLATFORM TO BE USED

- Learning is, according to its definition, information gathering process done through study. The specific case of machine learning, however, means that the learning task is performed by a computer, which in turn allows the building of programmes that constantly get better and better. These are the three fields of artificial intelligence applications.
- Data mining: Systems like these are designed to enhance power by making right decisions based on colossal amounts of data, which are incomprehensible to humans. For example, in the medical field, this will be a very good use because it will allow the setting up of medical knowledge based on patient data.

- Software applications: Nothing in the natural world can be artificially created by man, no matter how ridiculous it may seem. But the limitations can get less strict with the help of machine learning. Right now, such techniques are being used successfully in various fields like speech recognition, picture sorting, and self-drive vehicles, as evidenced by the project of Automatic Numbers Generation.
- Self-customizing programs: Even though the average person might not recognize it, nearly everyone is indirectly using this last specialty every day. Indeed, it is precisely this technology that creates those personalized news feeds that the users see when they surf the web according to their interests.

By using different labeled training instances, the algorithms can build broad target functions; when these functions are applied to a new dataset that has never before been used, the outcome is correctly predicted as expected. This is, in fact, the way artificial intelligence algorithms work. The training dataset is made of samples obtained through the label-instruction method, while the testing dataset consists of brand new, unused data. These types of applications are, therefore, dependent on a very strong and comprehensive training dataset since a bad train usually leads to disappointing predictions. The result that is presented in the GUI window is also the one that appears on the html page corresponding to the main system. The output page contains a “Load Data” button. After the “Like” button is pressed, a script that gets values from the dataset is executed. There is nothing more in the minds of artificial intelligence systems than the above-mentioned bases. In fact, one can think of a myriad of ways to implement a machine learning algorithm. Therefore, the selection of the best design requires a scrutiny that is often done through the use of a paper trail of different data.

6. Conclusion

The intended AI-based road accident prediction system is likely to be a reliable, effective, and adaptable solution for the detection of accident-prone locations through the use of images of traffic and machine-learning methodologies. The model will be recognizing and predicting the occurrence of accidents by correlating data like the weather, road conditions, traffic volume, and visual indications. The authorities will be equipped with real-time insights by the system that will be able to minimize accidents, enhance road safety design, and quicken the decision process. The application of SVM and other ML algorithms is expected to lead the platform to attain a high level of prediction accuracy and comprehensibility. In the end, this endeavor is valued as a contribution to the development of road safety, intelligent transport systems, and data-supported traffic management.

K-means clustering is an approach of unsupervised learning that is used in this very work for the untouched data; consequently, the findings are not separated into any group. Regression methods were also utilized in this work to reveal the reasons of traffic accidents with the help of huge amount of accident data. The analysis is done to identify the accident-related areas that happen at the same time and are then represented in the form of a graph. This greatly enhances our understanding of the accident scenarios and causes. And in the long term, this supports the Government in updating the traffic safety regulations to be in tune with various types and conditions of accidents.

References

1. Z. Zhang et al., “Deep Learning Advances in Vision-Based Traffic Accident Anticipation: A Comprehensive Review of Methods, Datasets, and Future Directions,” *arXiv preprint*, 2025.
2. J. Li, X. Wu, and H. Zhao, “Recent Advances in Traffic Accident Analysis and Prediction: A Comprehensive Review of Machine Learning Techniques,” *arXiv preprint*, 2024.

3. S. Yang, L. Chen, and W. Xu, "A Systematic Review on Road Accident Prediction," *International Journal of Bifurcation and Chaos*, vol. 35, no. 4, 2025.
4. M. Shaik, A. Rahman, and R. Kumar, "A Review on Neural Network Techniques for the Prediction of Road Traffic Accident Severity," *Asian Transport Studies*, vol. 8, no. 2, 2021.
5. T. Adewopo et al., "A Review on Action Recognition for Accident Detection in Smart City Transportation Systems," *Journal of Electrical Systems and Information Technology*, vol. 10, 2023.
6. Y. Fang, R. He, and J. Wang, "Vision-Based Traffic Accident Detection and Anticipation: A Survey," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, no. 2, 2024.
7. P. Chitraranjan, R. Singh, and S. Mathew, "Vision-Based Collision Warning Systems with Deep Learning: A Systematic Literature Review," *Journal of Imaging*, vol. 11, no. 1, 2025.
8. S. Manguri, A. Hasan, and D. Nuri, "A Review of Computer Vision–Based Traffic Controlling and Monitoring," *UHD Journal of Science and Technology*, vol. 7, no. 2, 2023.
9. A. Silva, F. Torres, and M. Gomez, "Machine Learning Applied to Road Safety Modeling," *Journal of Traffic and Transportation Engineering*, vol. 7, no. 4, 2020.
10. M. Khan, S. Ali, and R. Sheikh, "Analysis of Road Traffic Accident Using AI Techniques," *Open Journal of Safety Science and Technology*, vol. 15, 2025.
11. M. Toyoda, D. Yokoyama, J. Komiyama, and M. Itoh, "Big Data Analysis for Traffic Accident Patterns," in *Proc. IEEE Int. Conf. Big Data*, Boston, MA, pp. —, 2017.
12. J. R. Asor and G. M. B. Catedrilla, "A Study on Road Accidents Using Data Investigation and Visualization in Los Baños, Philippines," in *Proc. Int. Conf. Information and Communications Technology (ICOIACT)*, Yogyakarta, pp. 96–101, Mar. 2018.
13. S. Hussain, L. J. Muhammad, F. S. Ishaq, A. Yakubu, and I. A. Mohammed, "Performance Evaluation of Various Data Mining Algorithms on Road Traffic Accident Dataset," in *Information and Communication Technology for Intelligent Systems*, pp. 67–78, Dec. 2018.
14. T. B. Tesema, A. Abraham, and D. Ejigu, "Learning the Classification of Traffic Accident Types," in *Proc. Int. Conf. Intelligent Networking and Collaborative Systems*, pp. 463–468, Sept. 2012.
15. T. B. Tesema, D. Ejigu, and A. Abraham, "Knowledge Discovery from Road Traffic Accident Data in Ethiopia," in *Proc. World Congress on Information and Communication Technologies*, pp. 1241–1246, 2011.
16. G. Narasimhan, B. G. Ephrem, et al., "Predictive Analytics of Road Accidents in Oman Using Machine Learning Approach," in *Proc. Int. Conf. Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, pp. 1058–1065, July 2017.
17. L. Li, S. Shrestha, and G. Hu, "Analysis of Road Traffic Fatal Accidents Using Data Mining Techniques," in *Proc. IEEE Int. Conf. Software Engineering Research, Management and Applications (SERA)*, pp. 363–370, 2017.
18. N. R. and K. V., "Analysis of Road Accidents Using Data Mining Techniques," *International Journal of Engineering & Technology*, vol. 7, no. 3.10, pp. 40–44, 2018.
19. S. Hussain, "Survey on Current Trends and Techniques of Data Mining Research," *London Journal of Research in Computer Science and Technology*, vol. 17, no. 1, 2017.

20. J. Alcalá-Fdez, R. Alcalá, and F. Herrera, "A Fuzzy Association Rule-Based Classification Model for High-Dimensional Problems With Genetic Rule Selection and Lateral Tuning," *IEEE Transactions on Fuzzy Systems*, vol. 19, no. 5, Oct. 2011.
21. N. Donges, "The Random Forest Algorithm," *MachineLearning-Blog.com*, Feb. 2016.