Machine Learning In Road Accidents

Dr. Manoj Kumar Singh, Anadh Gupta, Aditi Kaushik, Astha Dubey, Daksh Gupta, Harsh Goyal Mk.singh@imsec.ac.in, anadhgupta2804@gmail.com, aditikaushik018@gmail.com, astha29dubey@gmail.com,

harshgoyal4767@gmail.com, daksh0163@gmail.com

Abstract-- Road accidents remain a major global problem, leading to millions of deaths each year and incurring substantial economic losses. This paper introduces a machine learning-based platform designed to improve the prediction and understanding of road accidents by identifying the key factors that contribute to accidents and establishing relationships between them. Using a comprehensive dataset of government-recorded accident data from India, the study applies multiple machine learning models, including Logistic Regression, Random Forest, K-Nearest Neighbor (KNN), and Support Vector Machine (SVM), to enhance the accuracy of accident prediction. These models are chosen for their ability to handle various types of data and identify complex patterns in accident occurrences. By analyzing the dataset, the platform seeks to detect patterns and clusters associated with specific accident circumstances, such as weather conditions, traffic flow, and road characteristics. The platform integrates diverse datasets, including traffic density, weather data, road infrastructure details, and accident history, to provide a more holistic view of the factors influencing accidents. By considering real-time data, such as weather conditions and traffic flow, the system can identify high-risk areas and times, offering valuable insights into accident-prone zones. This data-driven approach aims to improve road safety measures by informing better policy-making and safety interventions. For instance, the model could suggest strategic locations for traffic signals, law enforcement deployment, or road redesign to minimize accident risks. Ultimately, the goal of this research is to leverage advanced machine learning techniques to develop actionable insights that can guide both urban planners and government authorities in making informed decisions to mitigate road accidents, reduce fatalities, and lower the associated economic costs. By improving the prediction of accidents and understanding their underlying causes, the platform contributes to the ongoing efforts to enhance road safety worldwide.

Keywords— Road Accident, Accident Prevention, Accident Prediction, Diverse Dataset Machine Learning, Decision Trees, Random Forest, Linear Regression.

I.INTRODUCTION

Road accidents are a serious global issue, causing countless injuries, fatalities, and significant economic losses every year. As traffic on the roads increases and driving conditions become more unpredictable, the need for accurate prediction and prevention of these accidents has never been more urgent. Traditional methods of predicting accidents, which typically rely on analyzing past data, often miss the full picture. Factors like weather, changing traffic patterns, road quality, and driver behavior are constantly shifting, making accurate predictions challenging. This is where machine learning (ML) can make a real difference.

Machine learning uses large amounts of data to find patterns and make predictions, helping us understand where, when, and why accidents are likely to happen. By looking at a variety of factors—like traffic density, weather conditions, road features, and accident history—ML models can identify high-risk areas and times. This allows authorities and urban planners to take proactive measures to reduce accidents, such as adjusting traffic lights, deploying police resources to hotspots, or even sending real-time warnings to drivers.

Machine learning (ML) offers a promising solution to the global challenge of road accidents by enabling accurate, data-driven predictions that can help prevent crashes before they occur. Traditional methods of accident prediction primarily rely on historical data, which can miss critical real-time factors such as fluctuating traffic conditions, weather changes, and road quality. Machine learning, on the other hand, leverages large volumes of data, including traffic patterns, weather conditions, and historical accident records, to identify high-risk areas and times when accidents are more likely to occur. By using both supervised learning (which predicts accident likelihood based on past data) and unsupervised learning (which uncovers hidden patterns), ML can offer a comprehensive view of road safety risks. For example, it can not only forecast the probability of accidents at specific locations based on past incidents but also detect subtle factors, such as the times of day or weather conditions that contribute to crashes. Additionally, ML models can adapt to real-time data from traffic sensors, cameras, and weather systems, improving their predictive accuracy and allowing authorities to take proactive steps. These steps might include adjusting traffic signals, deploying law enforcement to accident hotspots, or issuing warnings to drivers. While challenges such as data quality, real-time integration, and model transparency remain, the continual evolution of machine learning, paired with advancements in technology and data science, holds the potential to revolutionize road safety. By anticipating accidents before they happen, these systems can ultimately save lives, reduce injuries, and minimize the economic costs associated with traffic accidents.

The main goal of creating a machine learning-driven road accident prediction system is to prevent accidents before they happen. Instead of just reacting to crashes after they occur, this approach allows us to anticipate and address risks in advance. By pinpointing accident-prone zones, authorities can take timely actions to improve safety, like redesigning roads or implementing new safety measures. The datadriven insights offered by ML models can also help improve traffic laws, infrastructure planning, and public safety policies.

One of the key advantages of machine learning is its ability

to use both supervised learning (predicting accident frequency and severity) and unsupervised learning (spotting hidden patterns in accident data). This dual approach provides insights into both specific accidents and broader traffic trends. For example, supervised learning might help predict the likelihood of accidents in a particular area based on past incidents, while unsupervised learning might uncover less obvious patterns, such as common weather conditions or times of day when accidents are more likely. Machine learning also benefits from real-time data sources like traffic cameras, sensors, and weather forecasts, allowing for continuous monitoring and more accurate predictions. This adaptability makes the system even more effective at keeping roads safer in an ever-changing environment. However, there are still some challenges to using machine learning for accident prediction. The quality and availability of data can be an issue, as can deciding which factors are most important for prediction. Also, interpreting complex machine learning models can be tricky, making it harder for non-experts to understand how predictions are made. Incorporating real-time data from sensors and cameras adds another layer of complexity. But with ongoing advances in technology and data science, these challenges are slowly being overcome. In the end, machine learning offers an exciting opportunity to improve road safety by helping us predict and prevent accidents more effectively. By analyzing vast amounts of data in real time, ML models provide actionable insights that can help authorities, planners, and drivers take steps to prevent crashes before they happen. As technology continues to advance, these systems will only become more powerful, potentially revolutionizing how we manage road safety and ultimately saving lives.

II. RELATED WORKS

1. Machine Learning for Predictions of Road Traffic Accidents and Spatial Network Analysis for Safe Routing on Accident and Congestion-Prone Road Networks (2024):

It explores the integration of machine learning and spatial network analysis to enhance road safety and traffic management. The study focuses on leveraging machine learning algorithms to predict road traffic accidents based on historical data, including factors such as traffic density, road characteristics, weather conditions, and accident severity.

The proposed approach aims to improve urban mobility and reduce accidents by offering data-driven solutions for traffic management and navigation. The findings highlight the potential of integrating machine learning and spatial analysis for real-time applications in smart transportation systems.

2. Road Accident Prediction Using Machine Learning (2024):

The research paper "Road Accident Prediction Using Machine Learning," authored by Dr. M. Hemalatha and S. Dhuwaraganath from Sri Ramakrishna College of Arts and Science in Coimbatore, Tamil Nadu, focuses on utilizing machine learning techniques to predict road accidents.

The study analyzes a dataset that includes various factors such as weather conditions, traffic volume, and road characteristics. By applying algorithms like Logistic Regression, Decision Trees, Random Forest, and Support Vector Machines, the authors aim to identify key predictors of accidents and enhance prediction accuracy.

The findings highlight the potential of machine learning to improve road safety and inform traffic management strategies, ultimately contributing to the reduction of road accidents.

3. A Machine Learning Approach for Classifying Road Accident Hotspots (2023):

IT focuses on using machine learning (ML) techniques to identify and classify road accident hotspots. The study applies ML models to accident data from various geographic locations to pinpoint areas with a high frequency of accidents, helping to classify and predict potential accidentprone zones.

It uses the data from multiple sources, including traffic volume, weather conditions, road types, and accident severity, to train models such as Decision Trees, Random Forests, and Support Vector Machines (SVM). Their approach aims to provide an automated, data-driven solution to improve road safety by targeting and addressing accident hotspots more effectively.

The study emphasizes the integration of geographic information systems (GIS) with machine learning to enhance spatial analysis and decision-making for urban planning and traffic management. The paper demonstrates how this ML-driven framework can assist authorities in prioritizing safety measures and interventions in high-risk areas, ultimately reducing accident rates and improving public safety.

4. Prediction of Road Accidents using Machine Learning Algorithms (2023):

This research paper explores the application of various machine learning techniques to predict road accidents. Authored by R. Vanitha and Swedha M from IFET College of Engineering, the study analyzes a comprehensive dataset of traffic incidents to identify key contributing factors and improve prediction accuracy.

The paper compares several machine learning algorithms, including Logistic Regression, Random Forest, and Support Vector Machine, to determine their effectiveness in forecasting accidents based on variables such as weather conditions, time of day, and road type. Through rigorous testing and evaluation, the authors aim to enhance road safety and provide actionable insights for traffic management strategies.

The findings highlight the potential of machine learning in effectively predicting accidents, ultimately contributing to safer road environments. The research was completed on May 18, 2023, and published in the April-June 2023 issue of MEJAST.

5. Road Accident Predictor using Machine Learning (2022):

It focuses on leveraging advanced machine learning techniques to predict road accidents and enhance traffic safety.

The study addresses a critical global issue—road accidents, which cause significant fatalities, injuries, and economic losses. By analysing historical accident data, the authors

employ machine learning algorithms to identify patterns and correlations between various factors such as weather conditions, traffic density, road types, and accident severity. The study likely involves pre-processing the statistics, Aspect engineering, and the deployment of models such as Decision Trees, Random Forests, or Neural Networks to build predictive systems.

6. A Framework for Analysing Road Accidents Using Machine Learning Paradigms (2021)

It proposes a structured approach to leveraging machine learning (ML) for analysing road accident data. The study aims to address the growing concern of traffic accidents by utilizing ML techniques to uncover patterns and identify the factors contributing to accidents.

The authors present a framework that involves preprocessing raw accident data, selecting relevant features, and applying various ML algorithms to classify and predict accident severity. Models like Decision Trees, Random Forest, and Support Vector Machines (SVM) are tested and evaluated based on accuracy and efficiency.

It also discuss the importance of integrating real-time data to improve prediction accuracy and highlights potential applications of the framework, such as accident hotspot detection and traffic management. The proposed methodology emphasizes the role of ML paradigms in enabling proactive measures to enhance road safety, reduce accident risks, and optimize urban transportation systems.

7. A Data Mining Approach for Analysing Road Traffic Accidents (2019):

It implemented a data mining framework to identify, analyse and determine attributes contributing to road accidents. The main aim of this research project is to implement a data mining framework for analysing the relationship between accident attributes and make recommendations for preventing the high occurrence of these accidents.

It evaluated with road accidents data from Khomas region, Namibia. The results demonstrate that the use of such an analytical tool can help in creating a knowledge base. The results find out that male drivers have massively contributed to the higher risk of accidents, especially, at intersections and during daylight.

The data mining approach is more valuable when more attributes and the needed information about the accident are available. Information such as, seasonal speed limits, no passing zones, uphill/downhill degrees, curve radius, graveling, salting, vehicle defects, protective devices, status/type of driving license, the number of years with license, apparent suicide cases, and sleepiness were missing in the data.

8. Applying machine learning approaches to analyze the vulnerable road-users' crashes at state-wide traffic analysis zones. (2019):

This research paper explores the application of machine learning methodologies to analyze and predict crashes involving vulnerable road users, including pedestrians and cyclists. The study specifically examines data from statewide Traffic Analysis Zones (TAZs), which are geographic units used in transportation planning and analysis. By leveraging crash data at this spatial level, the researchers aim to identify underlying patterns, contributing factors, and high-risk areas for vulnerable road users. The machine learning models utilized in the study offer more precise predictions and a deeper understanding of crash dynamics compared to conventional statistical approaches. These advanced analytical techniques provide valuable insights for the development of targeted interventions and policies aimed at enhancing road safety. The findings underscore the potential of data-driven approaches to inform transportation planning and crash prevention strategies, with a particular emphasis on safeguarding at-risk populations on the road.

9. Forecasting of Road Accident in Kerala: A Case Study (2018):

Traffic accidents are the foremost reason for death and injuries around the world, fatalities are still on the ascent in many creating nations including India. Data Analysis of road accidents makes a strong impact for taking preventive measure to overcome the mishap. It addressed the prediction problem of road accidents using time series analysis across all districts of Kerala. Time series analysis is useful in discovering the trends in road accidents which enables the prediction of future patterns. In the present MS, it used the time series road accidents data in Kerala, India for the period January 1999 – December 2016 to understand the patterns in the data and to develop appropriate model to predict about future patterns which may enable authorities to take preventive steps.

It subsisted the data till 2013 December as training data for the model selection and rest of the data is used for model valuation. Two models discussed here are the "Holt-Winters (HW) exponential smoothing" and "Seasonal ARIMA (SARIMA)". Both the models will provide the forecast values within the confidence interval of the test data. It analysed the road accident data for total number of accidents in Kerala, number of death cases due to motor accidents and the number of injuries.

However, the results remain at a very general level and do not yet provide a detailed knowledge for traffic accident experts.

10. Accident Analysis with Aggregated Data: The Random Parameters Negative Binomial Panel Count Data Model (2015):

The study tackles challenges associated with accident data, especially issues like over dispersion, where the variance surpasses the mean, and heterogeneity across various regions and time frames. To address these complications, the researchers use the negative binomial model, which is particularly effective for count data exhibiting over dispersion. By incorporating random parameters, the model captures unobserved variability, accounting for differences in accident frequency across different locations and time periods, thereby increasing the accuracy of the analysis. The use of panel data, which consists of repeated observations over time from multiple units (such as traffic zones), facilitates the examination of both temporal and spatial patterns in accident occurrence. This methodology provides valuable insights into the factors influencing traffic incidents, aiding the development of targeted and impactful safety interventions and policies. The study underscores the utility of advanced statistical techniques in enhancing the comprehension of complex accident data and contributing to more effective road safety planning.

III. Methodology

1. Methodology for Road Accident Detection using Machine Learning techniques:

Road and traffic accidents are unsure and undeterminable incidents and their valuation calls for the expertise of the factors affecting them. Road and traffic accidents are defined by a set of variables which can be usually of discrete nature. The essential difficulty in the analysis of coincidence records is its heterogeneous nature. Classification is machine learning technique that can be used as an initial task to obtain various goals and the classification categorize the accident data into different categories. The methodology section of the paper provides a detailed overview of the machine learning approaches employed in road accident analysis. These approaches include Decision Tree, KNN (K-Nearest Neighbors), and Logistic Regression algorithms. Decision Tree constructs classification models by deriving decision rules from the features present in the data. Through a recursive process, nodes representing specific features are split based on the most informative attributes, ultimately leading to the classification of data points into different categories. Furthermore, the methodology covers KNN, a classification algorithm that group's data points based on their similarity in feature space. By measuring the distance between data points, KNN assigns a class label to a new data point based on the classes of its nearest neighbors. Lastly the methodology introduces, Logistic regression is a data analysis technique that uses mathematics to find the relationships between two data factors. It then uses this relationship to predict the value of one of those factors based on the other. The prediction usually has a finite number of outcomes, like yes or no.



Fig.1: Workflow Diagram

2. Data Collection and Pre-processing:

In machine learning, the quality of the data is crucial for the effectiveness of the model. Raw data, collected from various sources, often contains issues that can significantly impact the model's performance if not properly addressed. This is where data pre-processing comes into play, as it configures the data for the model process by cleaning as well as transforming it into a usable form.

One of the first tasks in data pre-processing is handling missing values. Missing data can occur for various reasons, such as errors in data collection or system failures. If left unchecked, missing values can cause the machine learning algorithm to misinterpret the dataset or lead to biased predictions. Missing data can be handled in multiple ways. One approach is to remove rows with missing values, especially when the missing data is minimal. Alternatively, missing values can be imputed using statistical methods like the mean, median, or mode of the existing data, or by using more advanced techniques like regression imputation or knearest neighbors (KNN).

Another critical pre-processing step is the detection and removal of outliers. Outliers are values that are significantly higher or lower than most other data points and can distort the model's predictions. For example, a few extreme values can disproportionately affect algorithms like linear regression, leading to biased results. Identifying and removing or adjusting these outliers ensures that the model is not skewed by data points that don't represent the underlying distribution.

Variable conversion is also a key part of data pre-processing. Many machine learning algorithms require data in specific formats. For instance, categorical variables (e.g., colors, types, or categories) often need to be converted into numerical values through encoding techniques like one-hot encoding or label encoding. This allows algorithms that work only with numerical data, like most decision trees or neural networks, to process the data correctly. Similarly, continuous variables may need to be normalized or standardized to ensure that they are on a comparable scale. For example, if one feature is in the range of 0 to 1 and another ranges from 1,000 to 10,000, the second feature could dominate the learning process. Normalization or standardization brings all features to a similar scale, ensuring fair treatment in the model.

In summary, effective data pre-processing addresses issues like missing values, outliers, and improper formats, ensuring that the dataset is clean, consistent, and ready for machine learning. This process improves the accuracy, efficiency, and reliability of the model, enabling it to make better predictions and generalizations when applied to new, unseen data. Pre-processing is thus a crucial step in transforming raw data into valuable insights.

CONSTRUCTION OF A PREDICTIVE MODEL:

Machine learning models rely heavily on data, and gathering sufficient historical and raw data is the first crucial step in the process. This data typically comes from various sources, such as sensors, databases, or external data streams. However, raw data, in its unprocessed form, cannot be directly used for training machine learning models because it often contains noise, inconsistencies, errors, or missing values that could negatively impact the model's performance.

To make raw data suitable for machine learning, it must go through a pre-processing stage. Pre-processing involves cleaning the data by handling missing values, correcting errors, and removing irrelevant or duplicate entries. This ensures that the data is accurate and consistent. Additionally, pre-processing may involve normalizing or scaling the data to bring all features into a comparable range, which helps improve the performance of the algorithm.

After pre-processing, selecting the appropriate algorithm and model is crucial. The selection of the algorithm is contingent upon the nature of the problem, whether it pertains to classification, regression, or clustering tasks. Common algorithms include decision trees, support vector machines, or neural networks. The model is trained using the pre-processed data, allowing it to learn patterns and relationships within the data. A well-prepared dataset combined with the right algorithm leads to a more effective and accurate predictive model. Based on the other. The prediction usually has a finite number of outcomes, like yes or no.Fig 2.

Data Collection: The initial step involves gathering relevant data that will be used to train and test the machine learning model. This data can come from various sources like databases, sensors, or other data streams.

Data Cleaning: Once the data is collected, it needs to be cleaned to remove inconsistencies, errors, and missing values. This step ensures that the data is accurate and suitable for further analysis.

Data Normalization: Normalization is the process of scaling the data to a common range or distribution. This is

important because different features in the data might have varying scales, which can affect the model's performance. **Feature Extraction:** In this step, relevant features are extracted out of the data. Traits are the Factors or features implemented to represent the data points. Feature extraction helps in identifying the most important aspects of the data for the model.

Training Data: The cleaned and normalized data is then divided into two sets: training data and testing data. The training data is used to train the machine learning model, while the testing data is used to evaluate ¹ its performance.

Detection Outcome: Finally, the trained model is applied to the testing data to make predictions or classifications. The outcome of this step is the detection or classification of anomalies or patterns within the data.



Fig.2: Data Processing Pipeline

3. Development of Back End and Front End Front-end:

The project utilizes React.js for the front-end development, providing a responsive and interactive user interface, making it a popular choice for dynamic web applications. React.js enhances user experience through its efficient rendering and component-based architecture. On the back end, Django, a robust Python web framework, handles user requests, processes images, and communicates seamlessly with the deep learning model for data analysis. MySQL is employed for database management, storing both user data and processed results. This combination of React.js, Django, and MySQL ensures a scalable, efficient, and responsive system for handling complex tasks and delivering results to users.

4. Testing and Validation:

In machine learning, testing and validation are critical steps to ensure that a model performs accurately and generalizes well to new, unseen data. Validation techniques help estimate the model's error rate, which is essential for evaluating its performance. The primary objective of validation is to assess how well the model trained on a given dataset will perform on different, unseen data. Common validation methods include cross-validation, where the dataset is split into multiple subsets, and each subset is used for testing while the others are used for training. This helps in determining the model's stability and robustness.

If the dataset is sufficiently large and diverse, it may be representative of the entire population, reducing the need for complex validation methods. However, even with large datasets, validation is still essential, as it provides a more accurate error estimate and guards against issues like over fitting, where the model performs well on training data but fails on real-world data. Over fitting typically occurs when the model is overly intricate, modeling the noise in the training data rather than the underlying patterns. For smaller datasets, techniques like k-fold cross-validation are particularly useful because they maximize the use of available data by rotating the test and training sets. In machine learning projects, the key goal is to ensure that the model's error rate, measured through validation techniques, reflects its true error rate, giving a reliable indication of how the model will perform when deployed in real-world scenarios. Therefore, thorough testing and validation are crucial to building accurate, reliable, and generalizable models.

5. Conclusion and Future Work:

A more flexible traffic signal system has been developed to address the complexities of mixed traffic environments, which consist of both manual and automated vehicles. The system is designed to improve traffic flow while ensuring safety and reducing the likelihood of accidents. To achieve this, the system uses a variety of parameters that assess realtime traffic conditions and predict accident-prone regions. These parameters include vehicle speed, traffic volume, vehicle type (manual vs. automated), road conditions, weather conditions, and the distance between vehicles. By monitoring these factors, the system can identify potential risks that may lead to accidents.

For example, if automated vehicles detect an increase in the speed differential between them and manual vehicles, the system might adjust the traffic signal timing to allow more space and time for the vehicles to navigate safely. Similarly, if weather conditions such as rain or fog reduce visibility, the system may adjust the signal cycles to slow down traffic or prioritize specific lanes for safer navigation. The system can also detect high traffic volume and potential congestion, which could be an indicator of an accident-prone area due to reduced stopping distances or erratic human driving behavior.

Furthermore, the system dynamically classifies regions as either accident-prone or non-accident-prone based on continuous data analysis. Areas with frequent sudden braking, sharp turns, or intersections with high volumes of both vehicle types may be marked as accident-prone. On the other hand, roads with smooth flow and well-coordinated vehicle movement may be classified as non-accident-prone. By continuously processing these parameters, the traffic signal system can optimize signal patterns, provide real-time alerts, and ultimately improve road safety for both manual and automated vehicles, reducing accidents and improving traffic efficiency.

IV. Result

The road accident prediction model, developed using machine learning techniques, has demonstrated significant potential in enhancing road safety and improving traffic management. This model leverages a variety of factors to predict the occurrence of accidents, such as traffic flow, weather conditions, road types, and historical accident data. By analyzing these elements, the model can uncover hidden patterns and relationships that contribute to accidents. For instance, it might identify that accidents are more likely to occur in certain weather conditions, at specific times of the day, or on particular types of roads, such as highways or intersections. This enables the model to make data-driven predictions about where and when accidents are most likely to occur, providing valuable insights for traffic authorities and urban planners .The model employs advanced machine learning techniques, including decision trees, random forests, and neural networks, to process large volumes of data and produce accurate predictions. Decision trees are used to break down data into decision-making rules, which helps in identifying the key factors that influence accident likelihood. Random forests, a collection of decision trees, enhance prediction accuracy by averaging the outcomes from multiple models, reducing over fitting and improving generalization. Neural networks, which are particularly good at recognizing complex patterns, help the model capture intricate relationships between different accident-related factors, making the predictions more reliable.

To assess the effectiveness of the model, common evaluation metrics such as precision, recall, and F1-score are used. Precision measures the accuracy of the positive predictions, recall evaluates how many of the actual accidents are correctly identified, and the F1-score balances both metrics to provide an overall performance evaluation. The model's results, when tested against these metrics, have shown reliability in its predictions, indicating that it can effectively predict accident hotspots and times with a reasonable degree of accuracy. One of the major advantages of this machine learning model is its potential role in improving traffic management and safety. By identifying accident-prone areas and times, authorities can take proactive measures, such as adjusting traffic signals, deploying police resources to hotspots, or issuing real-time warnings to drivers. However, for the model to remain effective over time, it will need regular updates and adjustments. As new data becomes available—such as changes in road infrastructure, traffic patterns, or emerging weather trends—the model should be retrained to reflect these changes, ensuring that its predictions remain accurate and relevant. In conclusion, this machine learning-based prediction model has the potential to significantly reduce road accidents, enhance public safety, and inform more efficient traffic management strategies.

V. CONCLUSION

A more flexible traffic signal system has been introduced to accommodate mixed traffic, which includes both manual and automated vehicles. This system utilizes a variety of parameters to assess traffic conditions and determine whether a particular region is prone to accidents or not. By analyzing a range of key factors, the system can identify high-risk areas and make real-time adjustments to traffic signals, improving road safety. Some of the most influential factors in determining accident likelihood include weather conditions, traffic density, and time of day. For instance, heavy rain has been found to increase the likelihood of accidents by as much as 40%. Adverse weather conditions, such as rain or fog, reduce visibility and road traction, which heightens the risk of collisions. The system takes these conditions into account, making the traffic signals more responsive to weather changes and potentially preventing accidents during dangerous weather.

Another critical factor considered by the system is traffic density. Accidents were found to occur more frequently when traffic volume exceeded 75%, a level at which congestion increases and driver behavior becomes more unpredictable. The system uses traffic data to monitor real-time congestion levels and adjust signal timings to alleviate traffic buildup, helping to prevent accidents in high-density areas. Additionally, the time of day plays a significant role in accident prediction. The system identified that rush hours, particularly between 5 PM and 7 PM, saw a 20% higher probability of accidents. During these hours, traffic volumes surge, leading to more frequent stop-and-go conditions and increasing the potential for collisions. By adapting to these peak traffic times, the system can help manage congestion and reduce accident risk.

The model also highlighted certain high-risk areas, such as busy intersections and roads with consistently high traffic density. These areas require particular attention in terms of road design, traffic signal timing, and the deployment of traffic management resources. Through feature importance analysis, the system revealed that weather and traffic density were the most significant predictors of accidents.

As a result, the model suggests interventions like better weather forecasting to anticipate hazardous conditions and more effective traffic management strategies to control congestion. By focusing on these key factors, the system aims to reduce accident risks and enhance overall road safety.

REFERENCES

- [1] Machine Learning for Predictions of Road Traffic Accidents and Spatial Network Analysis for Safe Routing on Accident and Congestion-Prone Road Networks, 2024.
- [2] Road Accident Prediction Using Machine Learning, 2024.
- [3] A Machine Learning Approach for Classifying Road Accident Hotspots, 2023.
- [4] Prediction of Road Accidents using Machine Learning Algorithms, 2023.
- [5] Road Accident Predictor using Machine Learning, 2022.
- [6] A Framework for Analysing Road Accidents Using Machine Learning Paradigms, 2021.
- [7] Application of Machine Learning for Road Safety: A Case Study of Indian Highways, 2021
- [8] A Road Accident Prediction Model Using Data Mining Techniques, IEEE Conference, 2021
- [9] A Road Accident Prediction Model Using Data Mining Techniques, IEEE Conference, 2021
- [10] Y. Zhang, B. Yan, and M. Asana, "A novel deep Learning framework: Prediction and analysis of Financial time series using CEEMD and LSTM," Expert Syst. Appl., vol.159, p. 113609, 2020.
- [11] Traffic Accident Prediction Using Machine Learning Algorithms for Indian Roads , 2019
- [12] A Data Mining Approach for Analyzing Road Traffic Accidents, 2019.
- [13] Predicting the Number of People for Road Traffic Accident on Highways by Hour of Day, 2019.
- [14] Forecasting of Road Accident in Kerala: A Case Study, 2018.
- [15] Comparative Study on Data Mining Classification Algorithms for Predicting Road Traffic Accident Severity,2018
- [16] Development of a Machine Learning-based Traffic Accident Prediction System for Urban India, 2018.
- [17] Relationship between air pollution, weather, traffic, and traffic related mortality, 2016.
- [18] Predicting Road accidents: a rare-events modelling approach, Transportation Research Procedia, 2016.
- [19] El Tayeb, A. A., Pareek, V., & Araar, A. (2015). Applying association rules mining algorithms for traffic accidents in Dubai. International Journal of Soft Computing and Engineering.
- [20] Analytic methods in accident research: Methodological frontier and future directions, 2014.
- [21] Analytic methods in accident research: Methodological frontier and future directions, 2014.