IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

# Road Accident Prediction and Classification using Machine learning

**[1]Devansh Pradhan, [2]Apurv Upadhyay, [3]Ashish Gupta, [4]Arpita Pandey**

**[5]Dr. Rajkumar Gaur**

[1,2,3,4]Student [5]Guide
Department of Computer Science & Engineering,
Institute of Technology and Management, Gida, Gorakhpur, Uttar Pradesh, India

***Abstract:*** With the escalating number of vehicles, road safety has become a pressing concern, claiming 1.2 million lives annually worldwide. In Hyderabad alone, 2367 injury accidents were reported in 2017, causing significant loss of life and economic damage. To address this issue, we've employed machine learning algorithms to predict accident severity at specific times and locations. Factors such as speed limit, age, weather, vehicle type, light conditions, and day of the week were used for model training, leveraging a dataset from the UK government spanning 2005-2015. Random Forest was selected for its highest accuracy of 86.86%. Using Python, Scikit-Learn, NumPy, and Matplotlib, we built and tested the model on Google Collab and deployed it on Microsoft Azure's GPU-powered virtual machine. OpenWeatherMap API provides real-time weather and light data, while TextLocal API sends SMS notifications to authorities. A user-friendly web app, secured with HTTPS and a custom domain, facilitates input and output display. This predictive model promises to significantly aid traffic planning and management, potentially reducing road accidents in the future.

***Index Terms*** - Road Safety Prediction System, Accident Severity Prediction

## INTRODUCTION

The alarming prevalence of traffic accidents globally, as underscored by statistics from the World Health Organization (WHO), presents a dire challenge to public safety. Annually, these incidents claim the lives of 1.2 million individuals and leave 50 million others injured, resulting in staggering direct economic losses totaling $43 billion. With approximately 3,300 fatalities and 137,000 injuries occurring daily, the urgency to address this issue is paramount. Road accident prediction emerges as a crucial area of research in traffic safety, influenced by various factors such as road geometry, traffic flow patterns, driver behavior, and environmental conditions.

Numerous studies have been conducted to predict accident frequencies and analyze the characteristics of traffic accidents, including identifying hazardous locations, analyzing injury severities, and studying accident durations. While some research focuses on understanding the mechanisms of accidents, others explore the impact of weather and light conditions on road safety. Despite these efforts, there remains a lack of specific approaches available for traffic police to predict accident-prone areas at specific times. Such predictions play a crucial role in integrated traffic planning and management, considering the inherent randomness and nonlinear elements involved in traffic accidents, including human behavior, vehicle dynamics, road conditions, and climate.

Traditional linear analyses often fall short in accurately capturing the complex dynamics of traffic accidents due to limited data availability and the presence of noise pollution. Additionally, traditional backpropagation neural networks (BPNNs) suffer from various shortcomings, such as convergence to local minima, excessive iterations, and slow training processes. Consequently, there is a pressing need for advanced techniques that can overcome these limitations and provide more accurate predictions.

### 1.1 NEED OF THE STUDY

Machine Learning algorithms can process large number of classification parameters and are able to obtain useful patterns. It can process huge amounts of data efficiently and can be scalable. In computer science and related fields, artificial neural networks are computational models that simulate the central nervous system of the animal (especially the brain), allowing the machine to learn and identify information like the human brain.

### 1.2 PROBLEM DEFINITION

With the exponentially increasing number of vehicles, road safety is a matter of huge concern. Road accidents kill 1.2 million people every year. Road crashes cost $518 billion globally, costing individual countries from 1-2% of their economy. In 2017, there have been 2367 accidents with injuries reported in Hyderabad alone. Steps are being taken to combat this issue but they have been ineffective.

### 1.3 EXISTING SYSTEM

No specific approach available for the traffic police to predict which area is accident prone at a specific time. The traditional Back propagation network has defects. It has a 17% lower accuracy than the proposed model. We propose the use of a machine learning technique. Machine learning has the ability to model complex non-linear phenomenon.

### 1.4 PROPOSED SYSTEM

An ML powered web app which predicts accidents severity based on the current conditions. It is trained with 1.6 million accident records over 2005-2020. More data means greater accuracy. The purpose of such a model is to be able to predict which conditions will be more prone to accidents, and therefore take preventive measures. We will even try to locate more precisely future accidents in order to provide faster care and precaution service

According to the predicted severity, a message will be sent to the traffic police to take preventive measures

### RESEARCH METHODOLOGY

We have developed a web app for our model. It consists of four components:

**Front-End**: Users input for the prediction factors are taken and sent to the backend server.

**Back-End**: The model is deployed here and the input data is fed into the Machine Learning model.

**Machine Learning Model:** We have used decision tree, random forest and logistic regression and also applied hyperparameter tuning to increase its efficiency. Random Forest algorithm showed the highest accuracy of 86.86% and hence chosen for our model. The model runs and predicts the severity. The severity metrics are 1= Fatal, 2= Serious, 3= Slight. The output is sent back to the front-end and displayed to the user.

**Alert system:** An s.m.s containing the location coordinates and the severity of accident is sent to the police so that it can take preventive measures at the location.

## 2. REVIEW OF LITERATURE

In India, where over 150,000 lives are lost annually due to road accidents, and globally, nearly 50 million people are affected by roadside accidents each year, there is a pressing need for proactive measures to mitigate risks. This abstract introduces a machine learning-based solution utilizing random forest and Gaussian distribution techniques to identify and communicate potential road accident hotspots, referred to as "black spots," to users navigating unfamiliar locations. By employing predictive modelling, the system identifies regions with a higher likelihood of accidents, serving as a preventive tool to alert users through map notifications. The integration of this technology aims to reduce fatalities by providing users with critical information about potential risk areas, particularly beneficial for those traveling to unfamiliar locations where reliance on maps is common.

The proposed approach leverages the power of machine learning to offer proactive risk information, enabling users to make informed decisions while traveling. By combining random forest algorithms with Gaussian distribution methods, the system not only identifies possible accident-prone locations but also provides users with a visual representation of these areas on maps. This

innovation has the potential to significantly impact road safety by alerting users to high-risk zones, thereby contributing to a reduction in accidents and saving lives, especially in regions experiencing a rising trend in annual accident rates. [1].

The contemporary challenges in determining the root causes of traffic accidents, highlighting the intricate interplay of factors such as the driver's mental state, road conditions, weather, traffic density, and violations of traffic rules. Acknowledging the complexity, the paper emphasizes the growing role of machine learning classifiers in replacing traditional data mining methods like association rule mining. In particular, the study focuses on the application of four machine learning techniques—Naïve Bayes, k-Nearest Neighbours, Decision trees, and Support Vector Machines—to evaluate road accidents in Punjab. The research addresses the unique challenge of performing parametric evaluations tailored to the region, resulting in the identification of 12 highly relevant parameters. Notably, the Decision Tree classifier achieves the highest performance, reaching an accuracy of 86.25%, showcasing the effectiveness of machine learning in analysing and predicting road accidents in this specific context.

The study underscores the increasing popularity of machine learning applications in the realm of road accidents, providing a valuable insight into the performance of different classifiers in the context of Punjab. The identification of 12 pertinent parameters and the optimal accuracy achieved by the Decision Tree classifier signifies the potential of machine learning techniques to extract meaningful patterns and enhance our understanding of road safety dynamics, contributing to the development of targeted interventions for accident prevention in the region. [2]

This paper delves into the critical realm of road safety, acknowledging road accidents as a major global cause of injuries and fatalities. The focus of the research is on leveraging advanced algorithms, particularly machine learning techniques such as convolutional neural networks and long short-term memory networks, to analyse and predict traffic accidents. The overarching goal is to contribute to the creation of a safer mobility environment and ultimately save lives. The study comprehensively reviews the state of the art in road accident prediction, exploring various data sources including open data, measurement technologies, onboard equipment, and social media data. The classification of these sources based on origin and characteristics adds depth to the understanding of available information for predictive modelling.

The paper not only outlines the algorithms utilized for road accident prediction but also offers a comparative analysis of their effectiveness, considering factors such as ease of interpretation and the types of data being analyzed. Notably, the authors find that the most promising results arise from the combination of two or more analytic techniques, reinforcing the significance of a multifaceted approach to enhance predictive accuracy. Looking ahead, the paper identifies future challenges in road traffic forecasting, emphasizing the need to expand models by incorporating diverse data sources, including geo-spatial data, traffic volume, statistics, and sentiment from social media. This holistic approach is anticipated to further refine the precision and accuracy of predictive analyses, pointing towards a more comprehensive and effective strategy for accident prevention. [3]

This paper addresses the challenges in urban traffic forecasting by introducing a novel framework that integrates the strengths of Gaussian Mixture Model (GMM) and Support Vector Classifier (SVC). Traditionally, traffic forecasting models have leaned towards either GMM or SVC, each with its own advantages and limitations. While SVC can deliver robust performance with limited data, it comes at the cost of increased computational complexity. In response to this trade-off, the proposed framework combines the descriptive power of GMM with the high-performance classification capabilities of SVC. The innovative approach involves using the mean vectors derived from the GMM model as input for the SVC, creating a hybrid model that leverages the best of both techniques. Experimental results demonstrate the effectiveness of this approach, showcasing its superior performance compared to baseline statistical methods.

The hybrid framework presented in this paper not only overcomes the computational challenges associated with pure SVC models but also capitalizes on the descriptive richness of GMM. By integrating these two techniques, the approach achieves

favorable comparisons with baseline statistical methods in urban traffic forecasting. This suggests a promising avenue for developing more accurate and efficient models that can enhance our ability to predict and manage traffic patterns in urban environments.[4].

This study addresses the global concern of road traffic accidents (RTAs) by focusing on the analysis and modelling of accident data, aiming to enhance the understanding of the causes and effects of these incidents. Utilizing a real-life RTA dataset from Gauteng, South Africa, the research evaluates the performance of various machine learning classifiers. The classifiers considered include naïve Bayes, logistic regression, k-nearest neighbour, AdaBoost, support vector machine, random forest, and five missing data methods. The evaluation involves five key metrics: accuracy, root-mean-square error, precision, recall, and receiver operating characteristic curves. Additionally, the study incorporates parameter adjustment and dimensionality reduction techniques in its assessment. The empirical findings reveal that the random forest (RF) classifier, in conjunction with multiple imputations by chained equations, emerges as the most effective combination, outperforming other classifiers in terms of predictive performance.

The significance of this research lies in its practical implications for transport authorities and policymakers, providing insights into the optimal prediction model designs for RTAs. By systematically comparing various classifiers and employing comprehensive evaluation metrics, the study not only identifies the most successful combination (RF with multiple imputations) but also contributes to the broader discourse on leveraging machine learning for road safety analysis. This knowledge can inform evidence-based decision-making to improve road safety measures and reduce the impact of RTAs in the region and potentially serve as a valuable model for similar contexts globally.[5].

This paper addresses the complex issue of road accidents by exploring the multifaceted factors contributing to these incidents. Recognizing that driver emotions, environmental conditions, and various other variables can lead to accidents, the study focuses on extracting hidden patterns from road accident data. By applying machine learning techniques to the dataset collected from major national highways passing through Krishna district in 2013, the research aims to identify common features and patterns associated with accidents. The heterogeneous nature of the datasets is addressed through data cleaning measures, and relevant attributes are identified using attribute selection measures. The study employs K-medoids and expectation maximization algorithms to form clusters, and subsequently utilizes the a priori algorithm for pattern discovery. The results demonstrate the effectiveness of the selected machine learning techniques in uncovering hidden patterns within the data, providing valuable insights into the factors contributing to road accidents.

Furthermore, the paper contributes to the visual understanding of accident data by utilizing density histograms for visualization. This not only enhances the interpretability of the results but also offers a practical tool for stakeholders, policymakers, and researchers to comprehend and address the underlying patterns contributing to road accidents. By shedding light on the hidden relationships within the data, the research provides a foundation for informed interventions and strategies aimed at reducing the frequency and severity of road accidents on major national highways passing through the Krishna district.[6].

The escalating global concern over road safety is underscored by an alarming annual increase of over 4% in casualties across all age groups, with a forecasted growth of 8% in the casualty rate by 2030. Recognizing the urgency of addressing this issue, the paper emphasizes the need for in-depth analysis and proposes a comprehensive approach. Given the heterogeneous nature of road accident data, segmentation emerges as a critical task for effective analysis. The research advocates the use of the K-means clustering method to tackle this complexity, aiming to discern meaningful patterns and insights from the diverse data sets. By leveraging this segmentation technique, the model sets out to facilitate a nuanced understanding of the factors contributing to road accidents.

The second key task involves the extraction of data, images, and hidden patterns through the application of supervised machine learning algorithms. This step is pivotal for informing policies geared towards preventing road accidents. The combination of segmentation using the K-means clustering method and supervised machine learning contributes to generating meaningful information that can guide evidence-based policymaking. By adopting this holistic approach, the paper seeks to empower authorities and policymakers with valuable insights to formulate targeted strategies for the prevention of road accidents and the reduction of casualties.[7].

The surge in automobile numbers has led to a concerning rise in road accidents, causing non-fatal injuries and disabilities among victims. Addressing this issue, the paper focuses on classifying the severity of accidents by analyzing accident images, aiming to enhance road safety measures and expedite post-crash support services. The study utilizes feature extraction algorithms, including histogram of oriented gradient (HOG), local binary pattern (LBP), and speeded up robust features (SURF), to extract relevant features from accident images. These features are then inputted into k-nearest neighbour (KNN) and support vector machine (SVM) classifiers for severity classification. The comparative assessment of SVM and KNN classifiers, employing three different feature extraction algorithms, reveals that SVM outperforms KNN in accuracy. Specifically, SVM with HOG features demonstrates superior accuracy at 79.58% when compared to LBP and SURF, highlighting its effectiveness in classifying accident severity.

The research underscores the potential of image analysis and machine learning techniques in improving road safety. By successfully employing SVM with HOG features, the study not only contributes to the understanding of accident severity but also suggests a practical approach for real-world applications. The findings emphasize the importance of leveraging advanced algorithms for image analysis to enhance the efficiency of severity classification systems, ultimately aiding in the development of more effective post-crash support services and road safety strategies.[8]

This research addresses the intricate nature of determining the root causes of traffic accidents, emphasizing the contemporary challenges arising from a complex interplay of factors such as the driver's mental state, road conditions, weather, traffic dynamics, and adherence to traffic rules. The paper highlights the increasing popularity of machine learning classifiers as replacements for traditional data mining techniques like association rule mining in the analysis of road accidents. Focusing on the specific context of Punjab, the study employs four machine learning techniques—Naïve Bayes, k-Nearest Neighbours, Decision trees, and Support Vector Machines—to evaluate and understand road accidents in the region. The research faces the challenge of conducting parametric evaluations tailored to the unique characteristics of Punjab, aiming to extract highly relevant parameters.

The outcomes of the study provide valuable insights, revealing 12 parameters deemed most suitable for understanding and predicting road accidents in Punjab. Notably, the Decision Tree classifier achieves a maximum performance of 86.25%, showcasing the effectiveness of machine learning in this context. This research contributes to the ongoing efforts to improve road safety by identifying key parameters and leveraging machine learning techniques to enhance our understanding of factors influencing accidents, ultimately paving the way for more targeted and effective preventive measures in the region.[9]

This research paper delves into the substantial global issue of road crashes, where an alarming 1.3 million people lose their lives annually, with an average of 3,287 deaths per day and an additional 20-50 million sustaining injuries or disabilities. Road traffic crashes are identified as the 9th leading cause of global deaths, accounting for 2.2% of fatalities. The study focuses on addressing this pervasive problem through the application of soft computing, specifically leveraging fuzzy logic. The research centres on Greece, aiming to classify municipalities based on their respective occurrences of road accidents. To achieve this, a unique measure of similarity is employed, incorporating the concept of fuzzy transitive closure (FTC) to enhance the modelling of this complex and pervasive issue.

The contribution of soft computing, particularly fuzzy logic, to the modelling of road accidents in Greece is significant in providing a nuanced understanding of the varying occurrences across municipalities. By introducing a measure of similarity and utilizing fuzzy transitive closure, the research aims to contribute valuable insights that can inform targeted interventions and policies to mitigate the impact of road accidents. The application of soft computing techniques in this context showcases their potential to address complex, real-world problems and improve our ability to analyze and manage the multifaceted challenges associated with road safety.[10]

## REFERENCES

[1] : ALAGARSAMY, S., MALATHI, M., MANONMANI, M., SANATHANI, T., & KUMAR, A. S. (2021, DECEMBER). [2]: [1]: PREDICTION OF ROAD ACCIDENTS USING MACHINE LEARNING TECHNIQUE. IN 2021 5TH INTERNATIONAL CONFERENCE ON ELECTRONICS, COMMUNICATION AND AEROSPACE TECHNOLOGY (ICECA) (PP. 1695-1701). IEEE.

[2] : Singh, J., Singh, G., Singh, P., & Kaur, M. (2019). Evaluation and classification of road accidents using machine learning techniques. In *Emerging Research in Computing, Information, Communication and Applications: ERCICA 2018, Volume 1* (pp. 193-204). Springer Singapore.

[3] : Gutierrez-Osorio, C., & Pedraza, C. (2020). Modern data sources and techniques for analysis and forecast of road accidents: A review. *Journal of traffic and transportation engineering (English edition)*, *7*(4), 432-446.

[4] : Sangare, M., Gupta, S., Bouzefrane, S., Banerjee, S., & Muhlethaler, P. (2021). Exploring the forecasting approach for road accidents: Analytical measures with hybrid machine learning. *Expert Systems with Applications*, *167*, 113855..

[5] : Bokaba, T., Doorsamy, W., & Paul, B. S. (2022). Comparative study of machine learning classifiers for modelling road traffic accidents. *Applied Sciences*, *12*(2), 828.

[6]: VASAVI, S. (2018). EXTRACTING HIDDEN PATTERNS WITHIN ROAD ACCIDENT DATA USING MACHINE LEARNING TECHNIQUES. IN *INFORMATION AND COMMUNICATION TECHNOLOGY: PROCEEDINGS OF ICICT 2016* (PP. 13-22). SPRINGER SINGAPORE.

[7] : Yadav, J., Batra, K., & Goel, A. K. (2021, August). A Framework for Analyzing Road Accidents Using Machine Learning Paradigms. In *Journal of Physics: Conference Series* (Vol. 1950, No. 1, p. 012072). IOP Publishing.

[8] : Beryl Princess, P. J., Silas, S., & Rajsingh, E. B. (2021). Classification of road accidents using svm and knn. In *Advances in Artificial Intelligence and Data Engineering: Select Proceedings of AIDE 2019* (pp. 27-41). Springer Singapore.

[9] : Singh, J., Singh, G., Singh, P., & Kaur, M. (2019). Evaluation and classification of road accidents using machine learning techniques. In *Emerging Research in Computing, Information, Communication and Applications: ERCICA 2018, Volume 1* (pp. 193-204). Springer Singapore.

[10] : Katsoukis, A., Iliadis, L., Konguetsof, A., & Papadopoulos, B. (2018, July). Classification of road accidents using fuzzy techniques. In *2018 Innovations in Intelligent Systems and Applications (INISTA)* (pp. 1-275). IEEE.