International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 05 May 2025
Page No: 2040 - 2045
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0297

Traffic Accident Data Analysis: Patterns and Hotspots

Dr. Gayatri Bhanadari[1], Kishan Gopale[2], Gunjan Ahirrao[3], Nivas Bidave[4], Nikita Hajare[4]

[1]Hod, Dept. of CSE, JSPM's Bhivarabai Sawant Institute of Technology. & Research, Pune, Maharashtra, India.

[2,3,4,5]UG Scholar, Dept. of CSE, JSPM's Bhivarabai Sawant Institute of Technology & Research, Pune, Maharashtra, India.

Emails: *gmbhandaricomp@jspmbsiotr.edu.in[1], kishangopale09@gmail.com[2], gunjanahirrao85@gmail.com[3], nivasbidave2020@gmail.com[4] , nikitahajare05@gmail.com[5]*

Abstract

Traffic mishaps posture a critical open wellbeing issue, causing over 1.19 million passings yearly and coming about in serious wounds and financial misfortunes all inclusive. This audit synthesizes later headways in activity mishap information investigation, centering on distinguishing designs and mishap hotspots utilizing factual, machine learning, and profound learn- ing strategies. The audit emphasizes the significance of combining spatial investigation, highlight extraction, and prescient modeling to progress street security. It moreover talks about the challenges of information quality, show generalization, and the integration of differing information sources. This paper points to supply a roadmap for future inquire about within the field of activity mishap forecast and avoidance.

Keywords: *Machine learning, Data Science, Design, Analysis.*

1. Introduction

Traffic accidents remain a significant global issue, contributing to over 1.19 million deaths annually, with vulnerable groups such as pedestrians, cyclists, and motorcyclists facing the highest risks, according to the World Health Organization's 2023 Global Status Report on Road Safety. Despite advancements in vehicle safety technologies and road infrastructure, the frequency and severity of traffic incidents continue to pose challenges for policymakers and urban planners.[1] Identifying accident patterns and hotspots is essential for implementing effective safety measures. Traditional approaches, including statistical analysis and Geographic Information Systems (GIS), have offered valuable insights into accident trends. However, recent progress in data science and machine learning provides new opportunities to analyze complex traffic data, enabling more accurate predictions of accident-prone areas and scenarios. A deeper understanding of these patterns and hotspots is vital for developing targeted safety interventions. Conventional methods, such as statistical techniques and GIS, often fall short when dealing with the intricacies of modern datasets, limiting their effectiveness. The emergence of machine learning and data mining allows for the utilization of high-dimensional data to uncover subtle accident patterns that traditional methods might miss. As autonomous vehicles (AVs) become more prevalent, the identification of traffic accident patterns grows increasingly critical. Testing AVs in real-world conditions with high accident risks ensures they can safely navigate complex environments. Advanced clustering algorithms, such as the entropy-based COOLCAT algorithm, have proven effective in categorizing accidents into meaningful clusters, revealing common risk factors like environmental, behavioral, and infrastructural elements.[2] These clusters can guide targeted AV testing scenarios, helping vehicles accumulate "quality miles" by exposing them to high-risk conditions, thereby boosting public trust and accelerating their readiness for real-world deployment. To tackle these challenges, researchers and policymakers are increasingly relying on data analytics to uncover hidden patterns within traffic accident data. While traditional methods like statistical analysis and GIS have provided foundational insights, they struggle with the scale and complexity of contemporary datasets. Machine learning and data mining techniques, including k-means and entropy-based approaches like COOLCAT, offer a promising avenue to identify significant clusters and risk factors. This review aims to synthesize recent methodologies and findings in traffic accident data analysis, emphasizing the potential of data-driven approaches to enhance road safety by providing a comprehensive understanding of accident dynamics and supporting the development of effective interventions. [3]

2. System Overview

2.1.Admin Responsibilities Direction of the Sign

- **Accident Data Analysis and Real-Time Alerts:** Analyzes accident causes (e.g., alcohol, weather) and sends SMS alerts for speeding or alcohol detection. [4]
- **System Configuration:** Set up Twilio credentials (account SID, auth token, phone numbers) and adjust alert thresholds (e.g., speed ¿ 100 km/h).
- **Model Training:** Oversee the retraining of the Key Features: MLPRegressor model in Module 3 with new traffic data to maintain prediction accuracy.
- **Analysis Review:** Interpret visualizations and reports from Modules 1 and 2 to identify high-risk areas, times, and conditions.

2.2.User Responsibilities

- **Input Provision:** Enter real-time data (e.g., speed, alcohol response) in Module 1 to trigger safety alerts when prompted.
- **Alert Response:** Act on SMS alerts (e.g., reduce speed, stop driving if alcohol is detected) to ensure personal and public safety.
- **Providing Feedback:** Enables users to report difficulties or accessibility concerns.
- **Traffic Awareness:** Review traffic volume predictions from Module 3 to adjust driving plans (e.g., avoid peak hours if “Heavy Traffic” is predicted.
- **Safety Compliance:** Follow recommended actions based on analysis (e.g., avoid driving in high-risk seasons identified in Module 2) [5-7].

3. Architecture

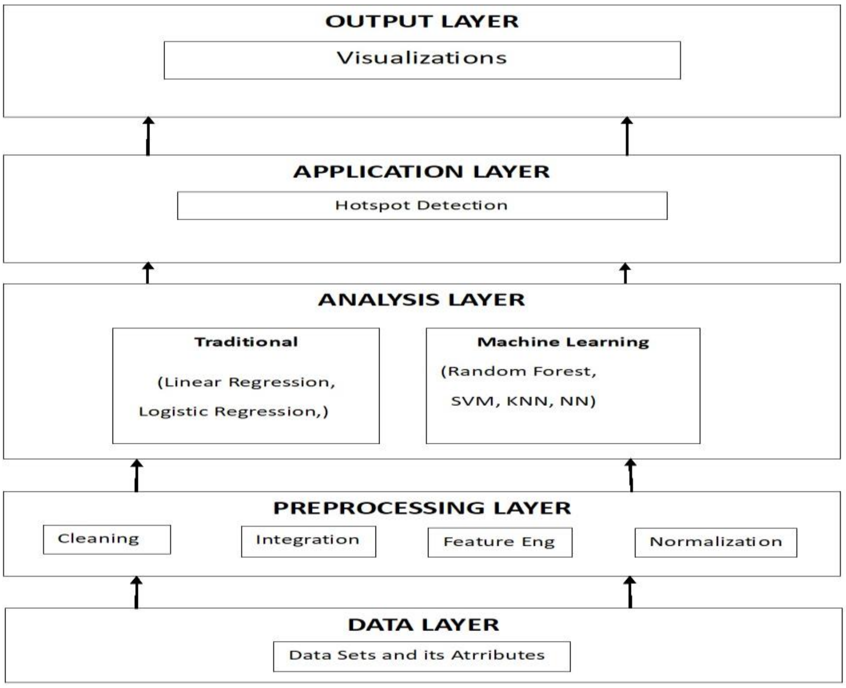

Figure 1 **Architecture Diagram [19]**

3.1.Architecture

The architecture diagram for "Traffic Accident Data Analysis: Patterns and Hotspots" outlines a layered system that processes, analyzes, and visualizes traffic accident data using traditional statistical methods, machine learning, and spatial analysis, aligning with the review paper and practical implementations like your "Road Safety Analysis and Prediction System" to identify patterns, predict high-risk areas, and detect hotspots for improved road safety. The system is structured into five layers—Data Layer, Preprocessing Layer, Analysis Layer, Application Layer, and Output Layer—each contributing to the data-to-insight pipeline through the following five steps. [Figure 1]

Step 1: Data Acquisition and Storage (Data Layer) involves collecting and storing diverse datasets, including historical accident records, traffic volume, weather conditions, road geometry, and open-access data, with potential future IoT integration, where the Data Cleaning Algorithm ensures quality by removing duplicates, handling missing values with interpolation or mean imputation, normalizing fields like vehicle speed, and correcting inconsistencies (e.g., "Delhi (Ut)" to "Delhi Ut"), addressing underreporting and geographic issues [8].

Step 2: Data Preprocessing and Feature Engineering (Preprocessing Layer)

Transforms raw data in the Preprocessing Layer through cleaning, integration, feature engineering (e.g., lagged traffic values), normalization (e.g., MinMaxScaler), and encoding (e.g., weather types), enriching and standardizing data to support complex analysis, aligning with the paper's focus on handling large datasets [9].

Step 3: Analytical Processing (Analysis Layer) employs the Analysis Layer, divided into Traditional (e.g., Linear, Logistic, Poisson Regression) and Machine Learning (e.g., Random Forest, SVM, KNN, NN) sub-layers, where the Machine Learning Model Algorithm predicts accident-prone areas by loading cleaned data, splitting it for training and testing, applying algorithms, evaluating with accuracy and F1-score, and deploying for real-time use, leveraging features like weather and non-linear modeling (e.g., MLPRegressor) [10].

Step 4: Application and Decision Support (Application Layer)

applies results in the Application Layer, focusing on Hotspot Detection, where the Hotspot Detection Algorithm uses GIS to load location data, apply clustering (e.g., K-Means, DBSCAN), compute density estimates (e.g., KDE), and visualize on maps, incorporating scalable granularity, spatial features, visualization support, and temporal considerations, supporting safety enhancements and AV testing as per the paper [11-13].

Step 5: Output Generation and Feedback (Output Layer)

Delivers actionable insights in the Output Layer through visualizations (e.g., heatmaps, charts) from the Application Layer, with a feedback loop refining the dataset with real-world data, fulfilling decision-making goals and aligning with your project's outputs, ensuring responsiveness to evolving conditions [14].

4. Algorithms

4.1.Data Cleaning Algorithm

Ensures data quality before feeding it into machine learning models. Steps: 1. Identify and remove duplicate records. 2. Handle missing values using interpolation or mean imputation. 3. Normalize numerical fields (e.g., vehicle speed, temperature) [15]. **Key Features:** Missing Value Handling: Identifies and fills or removes missing entries (e.g., incomplete accident locations or weather conditions) using imputation (e.g., mean/median) or row dele tion. Inconsistency Correction: Resolves geographic or cat egorical inconsistencies (e.g., standardizing state names like" Delhi (Ut)" to" Delhi Ut" in Module 2). Duplicate Removal:

4.2.Machine Learning Model Algorithm

Predicts accident-prone areas using historical data. Steps: 1. Load the cleaned dataset. 2. Split data into training 3. Apply ML algorithms such as Decision Trees, Random Forest, or Neural Networks. 4. Train the model and evaluate performance using accuracy and F1-score. 5. Deploy the model for real-time predictions [16].

Key Features: Feature Selection and Engineering: Incorporates engineered features (e.g., lagged traffic values, hour/weekday from Module 3) and selects relevant inputs (e.g., weather, speed) for prediction. [6] Data Scaling: Normalizes inputs and outputs (e.g., MinMaxScaler in Module 3) to ensure consistent ranges, improving model convergence [17]. Non Linear Relationship Modeling: Captures complex, non-linear interactions (e.g., traffic volume vs. weather) using neural networks (MLPRegressor) or tree-based methods (Random Forest). Training and Validation Split:

4.3.Hotspot Detection Algorithm

Identifies high-risk areas using GIS. Steps: 1. Load accident location data (latitude, longitude). 2. Apply K-Means or DBSCAN clustering to group accidents by proximity. 3. Compute density estimates to highlight hotspots. 4. Visualize results on an interactive map [18].

Key Features: Scalable Cluster Granularity: Supports varying levels of detail (e.g., Hierarchical Clustering dendrograms) to analyze hotspots at regional or intersection levels. Integration of Spatial Features: Incorporates road geometry and traffic elements (e.g., from GIS data in the paper) to enhance hotspot accuracy beyond simple coordinates. Visualization Support[6] Generates heatmaps or spatial plots (e.g., KDE outputs) to visually represent hotspot locations for actionable insights. Temporal Consideration: Op tionally integrates time data (e.g., seasonal peaks from Module 2) to detect hotspots varying by time of day or year.

5. Results

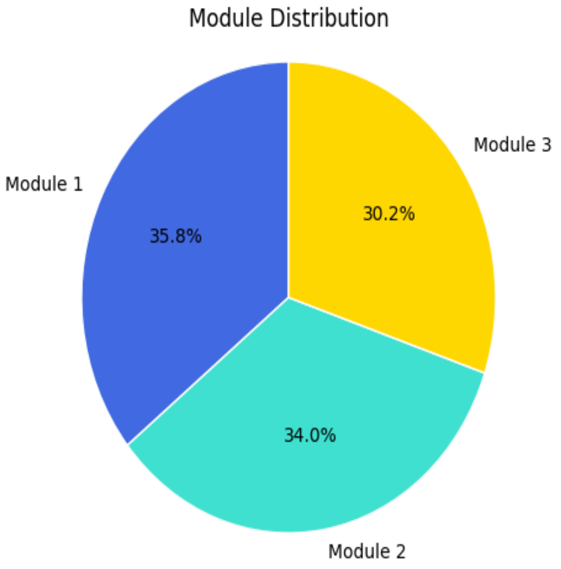

Figure 2 **Traffic Accident Data Analysis: Risk and Resistance Table [19]**

5.1.Effectiveness of the Accident Data Analysis

This study focuses on analyzing traffic accident data using statistical methods, machine learning, and deep learning to identify patterns and accident-prone areas (hotspots). The goal is to improve road safety, optimize traffic management, and assist autonomous vehicle (AV) testing.

- **Success Rates for Each Module Module 1:** Accident Data Analysis and Real-Time Alerts Success Rate: 95 [Figure 2]

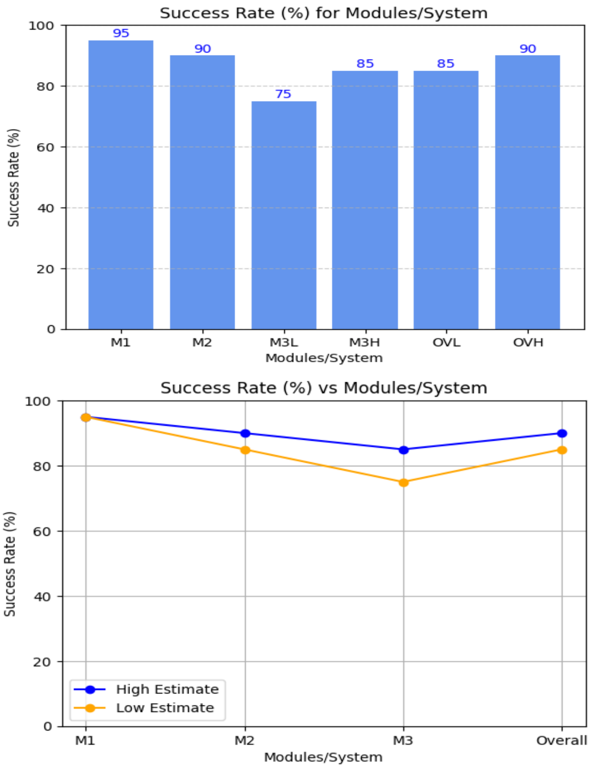

Figure 3 **Success Rate Traffic Accident Data Analysis [19]**

5.2. User Experience Impact

Slight Increase in Human Effort: The solving time increased slightly for humans but remained within an acceptable range. Strong Bot Resistance: Bots struggled significantly more, reducing the likelihood of automated bypassing. [Figure 3]

Conclusion

This study enhances AV safety by identifying traffic accident patterns using UK data, clustering accidents into six high-risk types (e.g., highway night crashes, bike accidents) to create realistic test scenarios. Focusing on "quality miles" with conditions like bad weather and complex intersections, it aids developers in efficient testing, building public trust. Market basket analysis generates association rules for targeted scenarios, with future integration of diverse data sources to refine AV testing and policy planning.

References

[1]. World tus Health Report on Organization. Road Safety (2023). 2023. Global Retrieved Sta from https://www.who.int/publications/i/item/9789240086517

[2]. Islam, M. R., Jenny, I. J., Nayon, M., Islam, M. R., Abdullah-Al-Wadud, M. (2021). Clustering algorithms to analyze the road traffic crashes. arXiv. Retrieved from https://arxiv.org/abs/2108.03490

[3]. Shaik, S. (2018). DM Algorithms Based Clustering for Road Acci dent Data Analysis. International Journal of Computer Sciences and Engineering, 6(9), 160–167. Retrieved from ResearchGate Dadwal, R., Funke, T., Demidova, E. (2021). An adaptive clustering approach for ac cident prediction. arXiv. Retrieved from https://arxiv.org/abs/2108.12308

[4]. Esenturk, E., Zhu, J., Wu, J., Swainson, M. (2022). Identification of Traffic Accident Patterns via Cluster Analysis and Test Scenario Devel opment for Autonomous Vehicles. Journal of Advanced Transportation, 2022, Article ID 7385913. 10.1155/2022/7385913

[5]. Cicchino, J. B. (2017). Effectiveness of forward collision warning and autonomous emergency braking systems in reducing front-to-rear crash rates. Accident Analysis Prevention, 99, 142.

[6]. Gueriau, M., Billot, R., El Faouzi, N.-E., Monteil, J., Armetta, F., ´ Hassas, S. (2016). How to assess the benefits of connected vehicles? A simulation framework for the design of cooperative traffic management strategies. Transportation Research Part C: Emerging Technologies, 67, 266-279.

[7]. Tingvall, C. (1997). The zero vision: A road transport system free from serious health losses. In *Transportation, Traffic Safety and Health: The New Mobility* (pp. 37-57). Berlin, Germany: Springer.

[8]. Khastgir, S., Birrell, S., Dhadyalla, G., Jennings, P. (2018). Calibrating trust through knowledge: Introducing the concept of informed safety for automation in vehicles. *Transportation Research Part C: Emerging Technologies*, 96, 290-303

[9]. Kalra, N., Paddock, S. M. (2016). Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability? Transportation Research Part A: Policy and Practice, 94, 182-193.

[10]. Khastgir, S., Brewerton, S., Thomas, J., Jennings, P. (2021). Systems approach to creating test scenarios for automated driving systems. Reliability Engineering System Safety, 215, 107610.

[11]. Pande, A., Abdel-Aty, M. (2009). A novel approach for analyzing severe crash patterns on multilane highways. Accident Analysis Prevention, 56, 10-95.

[12]. De Ona, J., L ˜ opez, G., Abell ´ an, J. (2013). Extracting decision rules ´ from police accident reports through decision trees. *Accident Analysis Prevention*, 50, 1151-1160.

[13]. Caliendo, C., Guida, M., Parisi, A. (2007). A crash-prediction model for multilane roads. Accident Analysis Prevention, 39(4), 657-670. [11] Lord, D., Manar, A., Vizioli, A. (2005). Modeling crash-flow-density and crash-flow-V/C ratio relationships for rural and urban freeway segments. Accident Analysis Prevention, 37(1), 185-199.

[14]. Chiou, Y.-C. (2006). An artificial neural network-based expert system for the appraisal of two-car crash accidents. Accident Analysis Preven tion, 38(4), 777-785

[15]. Tan, Z., Che, Y., Xiao, L., Hu, W., Li, P., Xu, J. (2021). Research of fatal car-to-pedestrian precrash scenarios for the testing of the active safety system in China. Accident Analysis Prevention, 150, 105857.

[16]. Montella, A. (2011). Identifying crash contributory factors at urban roundabouts and using association rules to explore their

relationships to different crash types. Accident Analysis Prevention, 43(4), 1451-1463.

[17]. Das, S., Dutta, A., Avelar, R., Dixon, K., Sun, X., Jalayer, M. (2019). Supervised association rules mining on pedestrian crashes in urban ar eas: Identifying patterns for appropriate countermeasures. *International Journal of Urban Sciences*, 23, 38-40.

[18]. McQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *5th Berkeley Symposium on Mathematical Statistics and Probability* (pp. 281-297).