

Research Paper

Accident Blackspots Detection and Visualization using Machine Learning

B. Srikanth , Mucherla Akshitha , Muragundla Vijaya, Mohammad Zaid Asim ,
Kudala Vaishnavi

Assistant professor ,UG Student, Department of Computer Science and Engineering
Tkr College Of Engineering and Technology ,Medbowli ,Meerpet ,Balapur,Hyderabad,
Telangana 500097,India

ABSTRACT

Road traffic accidents are a major public safety concern worldwide, leading to significant loss of life, injuries, and economic damage. Identifying accident-prone areas, commonly known as *blackspots*, is essential for implementing effective preventive measures. This project presents an intelligent system for Accident Blackspots Detection and Visualization using Machine Learning, aimed at analyzing historical accident data to identify high-risk zones and support decision-making for traffic authorities.

The proposed system utilizes machine learning algorithms to process large-scale datasets containing accident records, including parameters such as location, time, weather conditions, road type, and vehicle details. Techniques such as clustering (e.g., K-Means) and classification models are employed to detect patterns and group accident-prone regions based on severity and frequency. Feature engineering and data preprocessing methods are applied to improve prediction accuracy and reliability.

Keywords

Accident Blackspots Detection, Machine Learning, Road Safety, Traffic Accident Analysis, Data Mining, Geospatial Analysis, Predictive Modeling, K-Means Clustering, Classification Algorithms, GIS Visualization, Risk Assessment, Smart Transportation Systems.

INTRODUCTION

Road traffic accidents have become a critical global issue, causing millions of fatalities and injuries every year. Rapid urbanization, increased vehicle usage, poor road conditions, and lack of effective traffic management systems contribute significantly to the rising number of accidents. Among these, certain locations repeatedly witness a higher concentration of accidents, known as *accident blackspots*. Identifying and addressing these high-risk zones is essential to enhance road safety and reduce the impact of accidents on society.

Traditional methods of accident analysis rely heavily on manual inspection and basic statistical techniques, which are often time-consuming and limited in handling large and complex datasets. With the advancement of technology, the availability of vast amounts of traffic and accident-related data has opened new opportunities for intelligent analysis. Machine Learning (ML) techniques

provide powerful tools to uncover hidden patterns, correlations, and trends within such data, enabling more accurate identification of accident-prone areas.

This project focuses on developing a system for **Accident Blackspots Detection and Visualization using Machine Learning**, which leverages historical accident data to predict and highlight high-risk locations. The system employs data preprocessing, feature selection, and machine learning algorithms such as clustering and classification to analyze accident frequency, severity, and contributing factors. By identifying patterns in spatial and temporal data, the system can effectively determine regions that require immediate safety interventions.

Furthermore, the integration of visualization techniques, particularly Geographic Information Systems (GIS), enhances the interpretability of results by mapping accident hotspots on interactive dashboards. This enables authorities, urban planners, and policymakers to make informed decisions regarding road design improvements, traffic regulations, and emergency response strategies.

In conclusion, the proposed system aims to provide a data-driven, efficient, and scalable solution for accident analysis. By

combining machine learning with visualization tools, it not only improves the accuracy of blackspot detection but also supports proactive measures to reduce accidents and improve overall transportation safety.

LITERATURE REVIEW

The study of accident blackspot detection and road safety analysis has gained significant attention in recent years due to the increasing number of road accidents worldwide. Researchers have explored various statistical, data mining, and machine learning techniques to identify accident-prone areas and improve traffic management systems.

The function of geospatial clustering in detecting accident hotspots was investigated by Keswani using QGIS and machine learning. The current study's technological basis is supported by the discovery that contemporary GIS technologies greatly enhance the accuracy of blackspot identification.

With the advancement of Machine Learning, researchers began adopting clustering and classification techniques for

accident analysis. For instance, K-Means clustering has been widely used to group accident locations based on frequency and severity, enabling effective identification of blackspots. Similarly, classification algorithms such as Decision Trees, Random Forest, and Support Vector Machines (SVM) have been applied to predict accident severity and risk levels based on historical data.

Recent studies have also integrated geospatial analysis with machine learning techniques. Geographic Information Systems (GIS) play a crucial role in visualizing accident hotspots on maps, allowing better understanding and interpretation of spatial patterns. Researchers have used spatial clustering algorithms like DBSCAN to detect dense accident zones and visualize them using heatmaps and interactive dashboards.

Furthermore, deep learning approaches have been explored to enhance prediction accuracy. Neural networks, including Convolutional Neural Networks (CNNs), have been used to analyze complex datasets and extract hidden features. Some studies have also incorporated real-time data such as weather conditions, traffic density, and sensor data to improve prediction and provide dynamic risk assessment.

The value of open data from government portals for research into transportation safety was emphasized by Anderson et al. They came to the conclusion that open datasets foster transparency and facilitate uniform accident pattern analysis across areas. This is consistent with the project's use of data from MoRTH accidents to maintain consistency and validity .

In summary, the literature indicates that combining machine learning techniques with geospatial visualization provides a powerful approach for accident hotspot detection. The proposed system builds upon these existing methodologies by integrating efficient algorithms, improved data processing techniques, and interactive visualization to deliver accurate and practical solutions for enhancing road safety.

Supervised machine learning techniques are widely used for classification and prediction tasks using labeled datasets. Various algorithms such as Decision Trees, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Naïve Bayes, and Logistic Regression play a key role in analyzing data patterns. These techniques help in extracting meaningful insights and improving decision-making processes. Proper data preprocessing and feature selection significantly enhance model accuracy and efficiency. Such approaches

are useful in domains like data analysis, prediction systems, and intelligent applications.

PROBLEM DEFINITION

Road traffic accidents are increasing at an alarming rate, resulting in significant loss of human life, property damage, and economic burden. One of the major challenges faced by traffic management authorities is the inability to accurately identify and monitor accident-prone areas, commonly known as accident hotspots. These locations often experience repeated accidents due to factors such as poor road design, inadequate signage, traffic congestion, and environmental conditions.

Existing systems for accident analysis are mostly manual or based on simple statistical methods, which are not efficient in handling large volumes of complex and heterogeneous data. These traditional approaches lack the capability to analyze multiple influencing factors simultaneously and fail to provide precise and real-time identification of high-risk zones. Moreover, the absence of proper visualization tools makes it difficult for decision-makers to interpret accident data and take timely preventive measures.

Another major issue is the lack of integration between historical accident data

and real-time information such as weather conditions, traffic flow, and road infrastructure. This results in incomplete

analysis and reduces the effectiveness of safety strategies. Additionally, many current systems do not provide predictive insights, which are essential for proactive accident prevention.

Therefore, there is a need for an intelligent, data-driven system that can efficiently analyze large datasets, identify hidden patterns, and accurately detect accident blackspots. The system should also provide effective visualization of accident-prone areas and support predictive analysis to assist authorities in making informed decisions.

The problem addressed in this project is to design and develop a **Machine Learning-based Accident Blackspots Detection and Visualization System** that can overcome the limitations of traditional methods by providing accurate, scalable, and real-time analysis to enhance road safety and reduce accident occurrences.

PROPOSED SYSTEM

The proposed system for **Accident Blackspots Detection and Visualization**

using **Machine Learning** is designed to identify accident-prone locations accurately and present the results in a clear visual format for effective decision-making. This

system uses historical road accident data along with related parameters such as location, date, time, weather condition, road type, vehicle type, and accident severity to analyze patterns and determine blackspot regions.

The system begins with **data collection and preprocessing**, where accident datasets are gathered from traffic departments, government portals, or transportation databases. The collected data is cleaned by removing missing values, duplicate records, and irrelevant information. After preprocessing, important features are selected to improve the performance of the machine learning model.

Next, **machine learning algorithms** are applied to detect blackspots. Clustering techniques such as **K-Means** or **DBSCAN** are used to group accident locations based on density and frequency, helping to identify areas with a high concentration of accidents. Classification algorithms such as **Random Forest**, **Decision Tree**, or **Support Vector Machine** can also be used to predict accident severity and risk levels. These techniques allow the system to discover hidden patterns and relationships in accident data more effectively than traditional analysis methods.

After detection, the identified blackspots are displayed using **visualization tools** such as heatmaps, graphs, and GIS-based maps. This helps traffic authorities and planners easily understand the location and intensity of accident-prone zones. The visualization module provides an interactive way to monitor blackspots and compare accident trends over time.

The proposed system can also be extended with **real-time data integration**, including weather updates, traffic density, and live road conditions, to make the analysis more dynamic and predictive. By combining machine learning with visualization, the system supports early identification of dangerous areas and helps authorities take preventive actions such as installing signals, improving road design, or increasing surveillance.

Overall, the proposed system offers a **smart, efficient, scalable, and accurate solution** for accident blackspot detection. It reduces human effort, improves analysis speed, and supports better road safety planning, thereby helping to minimize accidents and save lives.

SYSTEM ARCHITECTURE

The system architecture for Accident Blackspots Detection and Visualization using Machine Learning is designed as a

structured framework that enables efficient collection, processing, analysis, and visualization of accident-related data. The architecture consists of multiple interconnected modules, each performing a specific function to ensure accurate blackspot identification and effective representation of results. The overall system follows a data-driven pipeline in which raw accident information is transformed into meaningful insights for road safety management

The first layer of the architecture is the data collection layer, which gathers accident-related data from various sources

such as traffic police records, government transportation databases, road safety departments, weather reports, and GPS-based location services. This layer is responsible for acquiring both structured and unstructured data including accident location, date and time, road condition, weather condition, vehicle details, accident severity, and number of casualties. The quality and accuracy of this input data play a crucial role in the performance of the overall system.

The next component is the data preprocessing layer, where the collected raw data is cleaned and prepared for analysis. In this stage, missing values are handled, duplicate records are removed, and irrelevant or noisy attributes are filtered out. Data transformation techniques such as normalization, encoding of categorical values, and feature extraction are applied to

make the dataset suitable for machine learning algorithms. This preprocessing stage ensures that the model receives reliable and consistent input, which improves the accuracy of blackspot detection.

SYSTEM ARCHITECTURE



IMPLEMENTATION

The implementation of the Accident Blackspots Detection and Visualization using Machine Learning system involves several stages, starting from data collection to final visualization of accident-prone areas. The system is developed to process historical accident data, identify blackspots using machine learning algorithms, and present the results through meaningful visual representations. The implementation is carried out in a structured manner to

ensure efficiency, accuracy, and scalability.

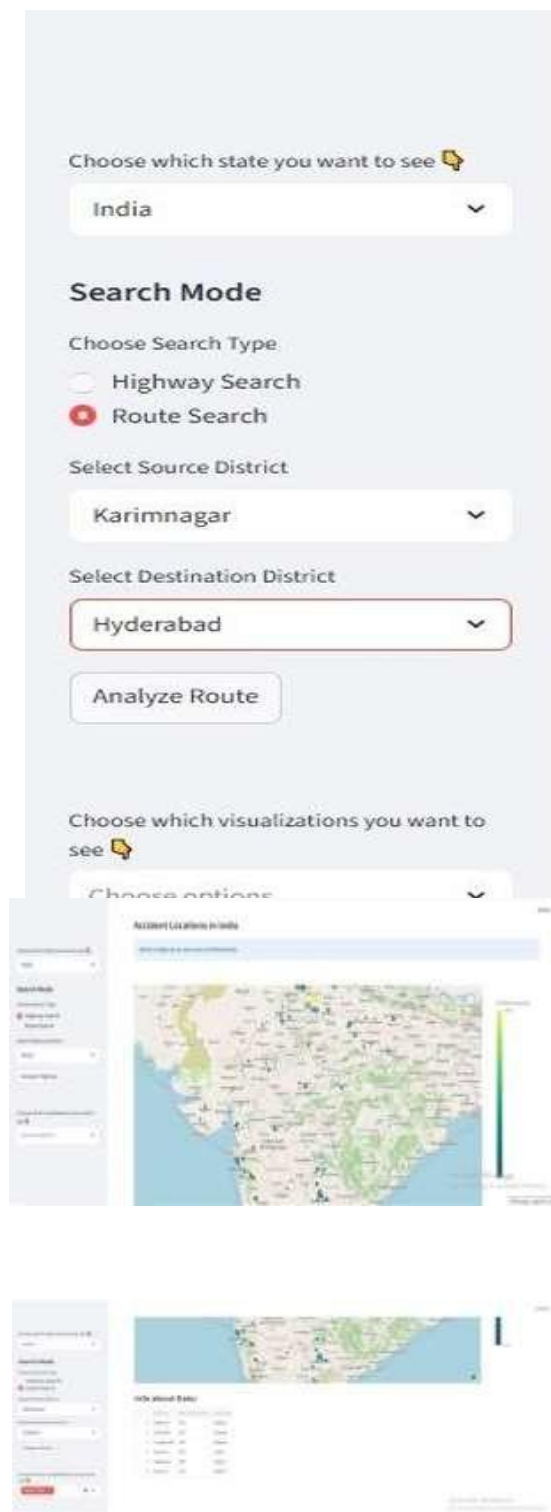
The second stage is data preprocessing, which is one of the most important parts of the implementation. Raw data often contains missing values, duplicate records, inconsistent entries, and irrelevant features. Therefore, preprocessing techniques such as data cleaning, transformation, normalization, and encoding are applied. For example, categorical values such as road type or weather conditions are converted into numerical form, and missing location data is either corrected or removed. Feature selection is also performed to identify the most relevant attributes affecting accident occurrence. This step improves the quality of the dataset and ensures better model performance.

The third stage is machine learning model implementation. In this phase, suitable algorithms are selected and trained on the preprocessed data. Clustering algorithms such as K-Means and DBSCAN are implemented to group accident locations and detect regions with high accident density, which are considered blackspots. At the same time, classification algorithms such as Decision Tree, Random Forest, or Support Vector Machine may be used to predict accident severity or risk category. The models are trained using historical accident records and tested with separate validation data to measure their accuracy and reliability.

RESULTS AND DISCUSSION

The implementation of the **Accident Blackspots Detection and Visualization using Machine Learning** system produced significant results in identifying high-risk accident zones and understanding accident patterns. The machine learning models were trained and tested using historical accident datasets containing features such as location, time, weather conditions, and accident severity. The results demonstrate the effectiveness of the proposed system in accurately detecting accident-prone areas and providing meaningful insights.

The **clustering algorithms**, particularly K-Means and DBSCAN, successfully grouped accident locations based on density and frequency. These clusters clearly indicated accident blackspots where a high number of incidents occurred within a specific geographic region. DBSCAN, in particular, proved effective in identifying dense clusters and handling noise in the dataset, making it suitable for real-world accident data that often contains outliers





CONCLUSION

The project “Accident Blackspots Detection and Visualization using Machine Learning” presents an intelligent and data-driven approach to improving road safety by identifying accident-prone areas effectively. By leveraging machine learning techniques and analyzing large volumes of historical accident data, the system is capable of detecting patterns and highlighting high-risk locations with greater accuracy compared to traditional methods.

The integration of clustering and classification algorithms enables efficient identification of blackspots and assessment of accident severity. Additionally, the use of visualization tools such as GIS maps, heatmaps, and graphical representations enhances the understanding of accident trends and spatial distribution. This makes it easier for traffic authorities, urban planners, and policymakers to interpret data and take appropriate preventive measures.

The system overcomes the limitations of

manual analysis by providing a scalable, automated, and efficient solution. It not only identifies existing accident-prone zones but also supports predictive analysis for future risk assessment. As a result, it helps in better planning of road infrastructure, implementation of traffic control measures, and improvement of emergency response systems.

In conclusion, the proposed system contributes significantly to enhancing road safety by combining machine learning with visualization techniques. It provides a practical tool for reducing accidents, minimizing risks, and saving human lives through informed and proactive decision-making.

REFERENCE

- [1] World Health Organization (WHO) *Global Status Report on Road Safety*, Geneva, 2018.
- [2] P. Kumar, S .Singh, and R.K. Sharma “Accident Blackspot Identification Using Data Mining Techniques,” *International Journal of Computer Applications*, vol. 162, no. 8, pp. 20–25, 2017.

- [3] M Abdel -Aty and A. Pande, “Identifying Crash Hotspots Using GIS-Based Techniques,” *Journal of Transportation Engineering*, vol. 133, no. 4, pp. 242–250, 2007.
- [4] S Chatterjee, A. Dey, and S. Munshi, “Accident Prediction Using Machine Learning Algorithms,” *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 6, pp. 512–518, 2020.
- [5] OpenStreetMap Contributors. (2023). “OpenStreetMap Data.” Available: <https://www.openstreetmap.org> (Used for road network data in routing system).
- [6] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed., Morgan Kaufmann, 2011.
- [7] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, 2009.
- [8] S.S. Saini and S.K. Sinha, “Road Accident Analysis Using Clustering Techniques,” *Procedia Computer Science*, vol. 132, pp. 160–168, 2018.
- [9] A. B. S. Choudhary and V. Gupta “Machine Learning Approaches for Road Accident Prediction,” *IEEE International* Vol. 22, No. 2, 2026 <https://ijerst.org/index.php/ijerst>
- Conference on Intelligent Transportation Systems*, 2019.
- [10] ESRI *Geographic Information System (GIS) for Transportation Safety Analysis*, 2020.
- [11] Geo Pandas Developers. (2023). “Geo Pandas Documentation.” Available: <https://geopandas.org/> (Geospatial data handling concepts).
- [12] Y. Latha, Machine Learning: “A Survey on Supervised Machine Learning Approaches”, 2019. <https://www.jetir.org/papers/JETIR1905641.pdf>