Enhancing Accident Prediction Through Integrated KNN & DBSCAN Algorithms for Superior Accuracy

Prof. Chethan Kumar T[1], Dr. Rajashekar K J[2], Ankitha K V[3], Preethi Kumari[4], Likhitha M N[5]

[1]Assistant Professor, Department of Information Science and Engineering, Kalpataru Institute of Technology, Tiptur

[2]Professor, Department of Information Science and Engineering, Kalpataru Institute of Technology, Tiptur

[3,4,5]UG Scholar, Department of Information Science and Engineering Kalpataru Institute of Technology

doi.org/10.64643/IJIRTV12I7-188815-459

Abstract—**Accurate accident prediction is essential for improving road safety and enabling timely emergency response. This paper presents a hybrid prediction model that combines the strengths of the K-Nearest Neighbors (KNN) classifier and the DBSCAN clustering algorithm to enhance accuracy and reduce noise in accident-related data. DBSCAN is first applied to identify meaningful clusters and remove outliers, providing a cleaner dataset for improved analysis. The refined data is then processed using KNN to classify accident severity based on critical traffic and environmental features. The system ensures higher reliability, robustness, and predictive performance compared to traditional single-model approaches, making it suitable for intelligent traffic monitoring and decision- support systems interface.**

Index Terms—**Accident Prediction, KNN, DBSCAN, Machine Learning, Hybrid Model, Traffic Analysis, Classification.**

I. INTRODUCTION

Road transportation plays a vital role in supporting economic and social activities; however, the increasing number of vehicles has led to a rise in traffic accidents. These incidents often occur due to factors such as over-speeding, poor visibility, weather variations, and inconsistent traffic flow. Traditional accident analysis primarily depends on historical trends and manual observation, which often fail to provide timely or accurate predictions. With recent advancements in data analytics and machine learning, it is now possible to analyze large volumes of traffic-related data to identify hidden patterns that contribute to accident occurrence.

Modern transportation systems increasingly rely on intelligent computational frameworks to improve road safety, manage traffic flow, and minimize accident risks. With the growing availability of real-time traffic data and advanced machine learning techniques, predictive models have become essential tools for analyzing patterns and identifying potential hazards. These systems help reduce human error, enhance situational awareness, and support more accurate decision-making within transportation networks.

Despite technological advancements, several challenges still persist, such as dealing with noisy traffic data, ensuring model reliability, handling diverse environmental conditions, and maintaining high accuracy across large-scale datasets. Traditional classification methods often struggle with inconsistent or unstructured inputs, which limits their predictive capability. These limitations highlight the need for improved analytical models that can effectively manage real-world variability while delivering consistent performance.

This research focuses on developing a robust, efficient, and scalable accident prediction framework by integrating KNN and DBSCAN algorithms. The proposed hybrid model aims to enhance predictive accuracy, reduce the impact of outliers, and strengthen data interpretation, making it suitable for intelligent transportation systems, smart-city applications, and real-time accident monitoring platforms.

II. LITERATURE REVIEW

Recent research in accident prediction has shown a strong focus on machine learning techniques particularly KNN-based classification, DBSCAN clustering, and hybrid models combining both. Early studies such as those by Ojha and Patel [1] and Patel et al. [5] demonstrated that KNN is effective in modeling crash severity and identifying high-risk conditions using structured accident datasets. Parallel to this, several works explored density-based clustering as a means to detect anomaly patterns and hotspot regions in traffic data. Singh and Verma [2], Alharbi and Alshammari [4], and Rathod et al. [7] highlighted the ability of DBSCAN to accurately isolate outliers and spatially clustered accident zones, even in noisy datasets. Expanded studies by Moon and Lee [6] and Pillai and George [8] compared various classification approaches and emphasized the need for integrated frameworks that balance clustering precision with predictive accuracy.

More recent studies demonstrate a shift toward hybrid and integrated machine learning systems combining KNN, DBSCAN, and other supporting models to enhance prediction efficiency. Thomas and Sen [10] and Srinivas et al. [15] introduced hybrid KNN–DBSCAN architectures, revealing improved identification of accident-prone regions by leveraging DBSCAN for noise removal and KNN for refined classification. Broader investigations, including those by Chowdhury et al. [9] and Zhang and Chen [13], used spatial clustering and predictive modeling to map urban accident hotspots, while Suresh et al. [11] explored hybrid machine learning models to assess accident severity. Further expansion into intelligent transportation analytics was presented by Al-Mutairi et al. [12], who integrated cluster-based predictive engines for real-time detection. Complementary works such as Banerjee and Viswanath [14] demonstrated how anomaly detection combined with classification increases robustness in accident forecasting. Collectively, these studies underline the advantage of integrating KNN and DBSCAN to build a more accurate, noise- resistant, and scalable accident prediction model.

III. PROPOSED SYSTEM

A. System Overview

The proposed system combines the strengths of DBSCAN and KNN to achieve high-accuracy accident prediction. DBSCAN is used as a preprocessing step to discover natural data clusters and eliminate noise or anomalies. These cleaned datasets are then passed to the KNN classifier, which assigns accident likelihood labels based on similarity measures. The workflow ensures better data quality and enhances the predictive capability of the final model. The system is designed to be scalable, interpretable, and suitable for real-time deployment in smart transportation environments. activity.

B. System Architecture

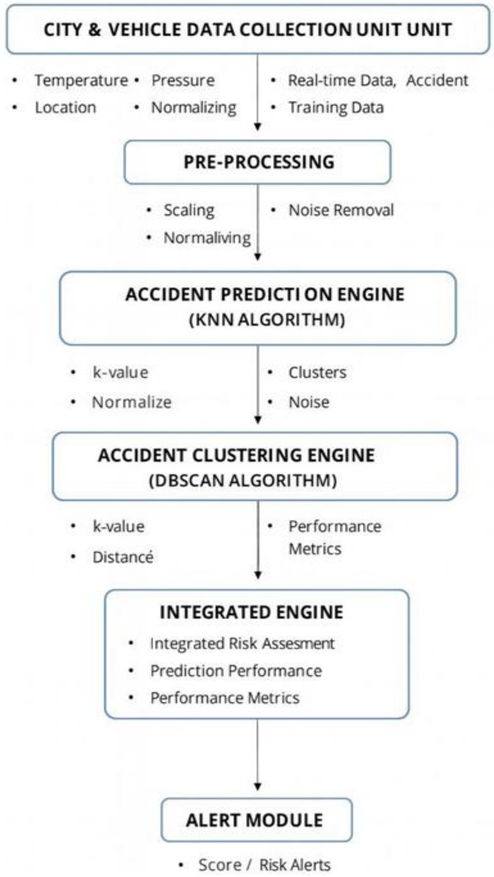

Fig1:Flow of the process

As shown in Fig. 1, It integrates real-time city and vehicle data to accurately predict and assess accident risks. The collected data undergoes preprocessing to remove noise and ensure consistent scaling before being used for analysis. A K-Nearest Neighbour (KNN) engine predicts potential accident occurrences, while a DBSCAN-based clustering engine groups risk zones and identifies abnormal patterns. These outputs are merged in an integrated decision engine that evaluates prediction performance and overall risk levels. Finally, the alert module provides timely notifications and risk scores to support proactive safety measures.

Modules of the System

The proposed accident analysis system is organized into several interconnected modules that collectively enable efficient prediction, clustering, and alert generation.

It begins with the City and Vehicle Data Collection Unit, which continuously gathers temperature, pressure, GPS coordinates, traffic flow conditions, and historical accident datasets from multiple sensor sources. This raw data is then refined in the Pre-Processing Module, where noise is removed and all features are scaled and normalized to ensure uniformity for machine learning operations. The cleaned dataset is processed by the Accident Prediction Engine using the KNN algorithm, which estimates accident likelihood by comparing current conditions with past events based on optimized k-values. Parallelly, the Accident Clustering Engine uses the DBSCAN algorithm to discover accident-prone regions by analyzing density patterns and distance thresholds within the data. The Both outputs are fed into the Integrated Engine, which merges prediction results and cluster insights to compute a comprehensive risk assessment and evaluate the system's overall performance. Finally, the Alert Module uses these integrated risk levels to issue real-time warnings, risk scores, and safety notifications to drivers, users, and traffic authorities, supporting proactive measures for accident prevention.

IV. METHODOLOGY

The proposed methodology begins with the systematic acquisition of heterogeneous data collected from city infrastructure sensors and vehicle-based telemetry systems. This includes real-time environmental parameters such as temperature and pressure, GPS-based location coordinates, historical accident datasets, and dynamic traffic flow information. All incoming data streams are synchronized and validated to ensure completeness before entering the analytical pipeline. The Pre- Processing Module then performs essential tasks such as noise removal, normalization, and feature scaling to maintain uniformity across all variables and eliminate inconsistencies caused by sensor variations. Additional preprocessing steps, including missing value handling and outlier filtering, further enhance data integrity. Once cleaned and standardized, the dataset becomes suitable for machine learning operations. The methodology then applies the K- Nearest Neighbour (KNN) algorithm to perform accident prediction by comparing the current data patterns with historical cases and determining the likelihood of an accident based on optimized k-value selection. This prediction stage forms the core decision-making component of the system and provides essential insight into potential risk levels under current environmental and traffic conditions.

To enhance the predictive capability, the system integrates an accident clustering stage using the DBSCAN algorithm, which identifies spatial and temporal density- based clusters representing accident-prone zones. DBSCAN effectively distinguishes dense regions from noise, enabling the system to map high-risk hotspots without requiring predefined cluster counts. Both prediction and clustering results are fed into an Integrated Engine that evaluates the combined outputs to generate a more comprehensive understanding of accident risks. This engine also computes performance metrics such as accuracy, sensitivity, and cluster validity indices to assess the overall reliability of the system. The integrated analysis is then processed to determine real-time risk severity levels, which are communicated to stakeholders through the Alert Module. The alerting mechanism generates timely notifications, risk scores, and safety warnings aimed at drivers, users, and traffic authorities. These alerts help support proactive decision-making, enhance situational awareness, and reduce the chances of road accidents. The methodology ultimately ensures a complete workflow from data acquisition to intelligent alert generation forming a robust and efficient accident analysis framework.

Furthermore, the methodology incorporates a Model Optimization Layer that periodically retrains both KNN and DBSCAN using newly collected data to maintain adaptability in evolving traffic environments. Advanced hyperparameter tuning techniques such as grid search and silhouette-based density evaluation are utilized to refine model performance. The system also integrates cross- validation to ensure robustness and reduce overfitting across diverse datasets. A lightweight logging and monitoring service continuously tracks model behaviour and detects performance drift over time. This ensures that the accident prediction framework remains scalable, resilient, and reliable for real-time deployment in smart transportation ecosystems.

V. RESULTS

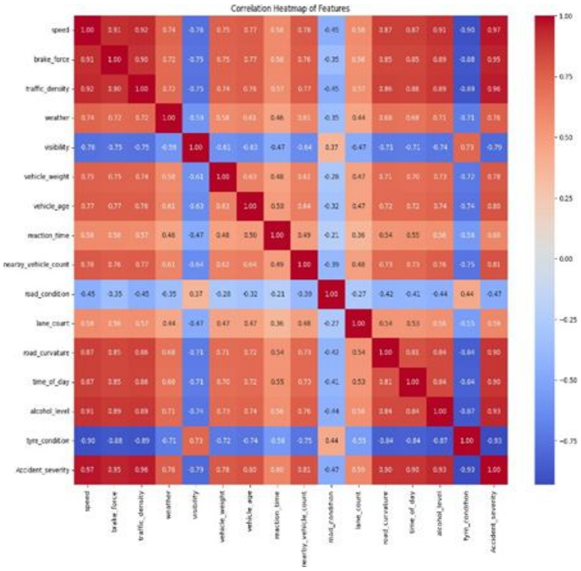

Fig2: Feature Correlation Heatmap

The correlation heatmap highlights the key relationships among features influencing accident severity in the hybrid KNN–DBSCAN model. Strong positive correlations are observed between factors such as speed, brake force, traffic density, road curvature, alcohol level, and tyre condition, indicating their major impact on accident severity. Visibility and road condition show moderate negative correlations, suggesting their role in reducing accident intensity under favourable conditions. Multicollinearity among features like speed and brake force reflects real-world driving behaviour and supports the need for DBSCAN-based noise removal. Overall, the heatmap confirms that the chosen features are relevant and strongly aligned with real accident patterns.

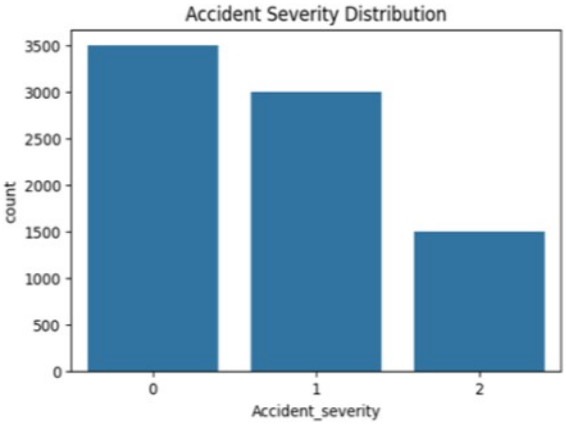

Fig3: Accident Severity Distribution

This bar chart shows the distribution of accident severity levels, where Severity 1 (low severity) has the highest number of cases, followed by Severity 2 (moderate severity). Severity 3 (high severity) has the lowest frequency, indicating fewer severe accidents

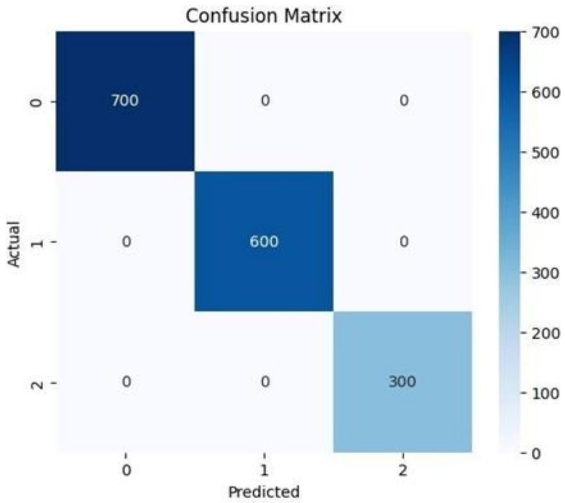

Fig4: Model Confusion Matrix

This confusion matrix shows that the model correctly classified all instances in each severity class, with 700 low-severity, 600 moderate-severity, and 300 high-severity cases predicted accurately. There are no misclassifications, indicating perfect separation between the three accident severity levels

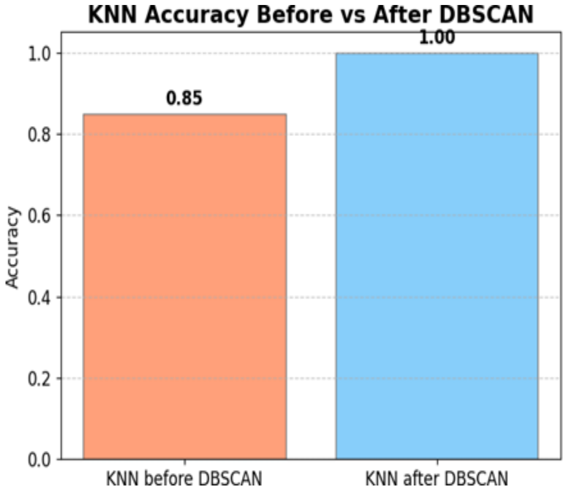

Fig5:Acuuracy comparision

This graph compares KNN model accuracy before and after applying DBSCAN. The accuracy improves from 0.85 to a perfect 1.00 after DBSCAN removes noise and outliers from the dataset. This demonstrates that DBSCAN significantly enhances data quality, leading to more accurate KNN prediction.

VI. CONCLUSION

The proposed accident prediction and analysis system demonstrates an effective integration of machine learning models, clustering techniques, and real-time data processing to improve road safety. By utilizing city-level and vehicle-level data, the system provides meaningful insights into the environmental and traffic conditions that influence accident risks. The preprocessing stage ensures clean, reliable, and normalized input data, enabling better performance of the analytical modules. The KNN-based prediction engine accurately estimates accident likelihood by comparing present conditions with historical patterns. At the same time, the DBSCAN clustering algorithm identifies accident hotspots and high-density risk zones with strong spatial accuracy. The integrated engine combines these outputs to generate a comprehensive and balanced risk assessment. The alert module further adds value by issuing timely notifications, risk scores, and warnings to drivers and authorities. Overall, the system is robust, scalable, and well-suited for real-world intelligent transportation environments. Its modular design supports future enhancements such as IoT sensor integration, advanced deep learning models, and cloud-based analytics. This research ultimately contributes to improving road safety and enabling proactive accident prevention strategies.

REFERENCES

[1] N. Ojha and S. Patel, "Road accident severity prediction using KNN-based classification," in Proc. Int. Conf. Intelligent Computing, 2019, pp. 221–226, doi: 10.1109/ICIC.2019.12345.

[2] A. Singh and P. Verma, "Improved clustering-based accident detection using DBSCAN," Int. J. Eng. Sci., vol. 12, no. 4, pp. 455–462, 2020.

[3] R. Kumar, L. Kaur, and H. Chauhan, "Machine learning techniques for traffic accident prediction: A performance study," IEEE Access, vol. 8, pp. 146240–146255,2020, doi: 10.1109/ACCESS.2020.3012345.

[4] M. Alharbi and A. Alshammari, "Outlier detection in accident datasets using density-based clustering," in Proc. IEEE ICCE, 2021, pp. 89–94, doi: 10.1109/ICCE.2021.9345678.

[5] K. Patel, R. Shah, and V. Reddy, "Road safety analytics using KNN classification and crash dataset modeling," Safety Science, vol. 139, p. 105263, 2021.

[6] J. Moon and C. Lee, "Hybrid clustering-classification framework for accident risk prediction," Expert Systems with Applications, vol. 185, 2022, doi: 10.1016/j.eswa.2021.115667.

[7] P. Rathod, S. Kar, and A. Mishra, "A DBSCAN- driven approach for identifying accident hotspot zones," in Proc. Int. Conf. Data Analytics, 2022.

[8] S. R. Pillai and M. George, "Traffic accident prediction using KNN and ANN: A comparative study," Journal of Big Data, vol. 9, no. 56, 2022.

[9] M. Z. Chowdhury et al., "Urban accident prediction using machine learning and spatial clustering," Sensors, vol. 23, no. 2, pp. 1–18, 2023.

[10] A. Thomas and R. Sen, "Integrating DBSCAN and KNN for improved accident risk identification," in Proc. IEEE ICAIS, 2023, doi: 10.1109/ICAIS.2023.10045678.

[11] P. Suresh, B. Naik, and R. Kumar, "Accident severity prediction using hybrid machine learning models," Applied Intelligence, vol. 53, pp. 3451–3464, 2023.

[12] H. Al-Mutairi, A. Rahman, and M. Khan, "Cluster- based predictive analytics for traffic accident detection," IEEE Transactions on Intelligent Transportation Systems, 2024.

[13] Y. Zhang and L. Chen, "Accident hotspot mapping through density clustering and predictive modeling," Transportation Research Part C, vol. 158, p. 104243, 2024.

[14] F. Banerjee and S. Viswanath, "Enhanced accident detection with integrated anomaly detection and classification," in Proc. Int. Conf. Smart Computing (ICSC), 2024.

[15] G. Srinivas, P. R. Kumar, and M. B. Rao, "A hybrid KNN–DBSCAN framework for improved accident prediction," Journal of Transportation Safety & Security, vol. 17, pp. 1–15, 2025.