Modeling the Accident Severity Level: A Survey

Lambodar Gourish Hebbar[1], Raghavendra Bhardwaj D G [2], Vardhan Kumar G K[3], Varun S[4], Roopashree C S [5],
[1,2,3,4] *Student, Department of ISE, BNM Institute of Technology, Bangalore, Karnataka, India*
[5] *Assistant professor, Department of ISE, BNM Institute of Technology, Bangalore, Karnataka, India*

***Abstract*—Road accidents are a significant problem that can cause minor to very serious injuries, even death. In this study, we use machine learning algorithms to predict how bad accidents might be and where they might happen. We looked at data from accidents in the UK and used it to figure out how severe each accident was. The outcome is classified as the severity of the accident on a scale of 1-5, where 1 is the least severe and 5 is the most severe. The results have been visualized to the user through an application. Also, the location of various accident-prone areas has been plotted on the map.**

I. INTRODUCTION

Road traffic accidents continue to pose a major global challenge, particularly in densely populated regions where high vehicle density and complex traffic behaviors increase the likelihood of collisions. These incidents not only lead to the tragic loss of human life and severe injuries but also place a heavy financial burden on individuals, families, and national economies. Despite numerous efforts to improve road safety, the number of accidents remains alarmingly high, highlighting the urgent need for more intelligent, data-driven solutions.

This project focuses on predicting the severity of road accidents based on various influencing factors, with the ultimate goal of enhancing public safety. By leveraging machine learning techniques and integrating diverse datasets including past accident reports, road infrastructure details, and traffic flow patterns we aim to identify high-risk areas where accidents are more likely to occur. Such insights can support authorities and urban planners in taking timely, targeted actions to reduce accident rates.

In addition, the project incorporates Google Maps APIs in Python to visualize accident-prone zones, offering a more intuitive and accessible understanding of potential hazards on the road. The development of a severity-ranking system will not only help prioritize safety interventions but also pave the way for future applications, but also pave the such as real-time alerts through a
mobile app. This approach bridges the gap between technology and public safety, fostering safer transportation environments through proactive and informed decision-making.

II. LITERATURE SURVEY

Over the past decade, the use of machine learning (ML) in road safety analysis has gained significant traction. Researchers have explored various models and data sources to better understand the factors contributing to road accidents and to develop predictive systems that can support decision-making and reduce accident risks.

In one notable study, Support Vector Machines (SVM), Decision Trees, and Logistic Regression were employed to classify the severity of accidents based on factors like weather conditions, road surface, and time of the day. The results showed that machine learning techniques could outperform traditional statistical models in predicting accident outcomes due to their ability to handle complex, nonlinear relationships in data.

Another study incorporated Geographic Information Systems (GIS) to spatially map accident hotspots, enhancing the visualization of critical areas. By combining spatial data with traffic density and road type information, researchers were able to identify zones that require immediate attention for infrastructure improvement or increased monitoring. Deep learning approaches have also emerged in this field. Convolutional Neural Networks (CNNs), though primarily used for image processing, have been adapted for analyzing structured tabular data when combined with spatial or visual features. These models demonstrated improved accuracy in predicting accident severity levels, especially when large-scale datasets were available.

Furthermore, the integration of real-time data such as traffic flow, weather updates, and vehicular speed from IoT devices has been explored in smart city projects. These dynamic data sources have proven valuable in improving the accuracy of accident prediction models and enabling timely intervention.

Studies have also emphasized the importance of historical accident records. Datasets such as the UK Road Safety Data and the US National Highway

Traffic Safety Administration (NHTSA) database have been widely used for model training and evaluation. These datasets provide structured records that include attributes like vehicle type, driver behavior, location, and casualty details, all of which are crucial for robust prediction.

In terms of user interaction, recent works have proposed the use of mobile applications and web dashboards to deliver insights to the public and local authorities. By integrating mapping services like Google Maps APIs, these platforms allow users to easily visualize accident-prone areas and access location-based safety recommendations.

Collectively, the literature highlights the potential of combining machine learning algorithms, spatial visualization tools, and multi-source data to build intelligent systems for accident severity prediction. However, challenges such as data imbalance, lack of real-time data access, and regional variability in traffic rules remain, suggesting areas for future research and innovation.

III. CHALLENGES

Despite the promising potential of using machine learning for accident severity prediction, several challenges need to be addressed to ensure the effectiveness and reliability of such systems.

1. Data Quality and Availability: One of the major hurdles in developing predictive models is the inconsistency and incompleteness of accident-related data. In many regions, accident records are either poorly maintained, lack essential details, or are not digitized. Missing values, inaccurate entries, and unbalanced datasets (with fewer severe accident cases compared to minor ones) can significantly affect model performance.
2. Imbalanced Datasets: Accident severity data often suffer from class imbalance, where minor accidents greatly outnumber fatal ones. This skewed distribution can cause machine learning models to be biased towards the majority class, leading to poor accuracy in predicting severe accidents—which are of greater concern.
3. Feature Selection and Relevance: Accident severity is influenced by a wide range of factors including weather, road type, lighting conditions, vehicle speed, and driver behavior. Selecting the most relevant features while filtering out noise is a complex task that requires domain expertise and thorough data preprocessing.
4. Real-Time Data Integration: Integrating real-time data such as live traffic, weather updates, and vehicle telemetry can enhance prediction accuracy, but acquiring and processing such data in real-time is technically demanding. It often requires APIs, IoT devices, and infrastructure support, which may not always be accessible.
5. Regional Variability: Traffic patterns, road conditions, and driving behavior vary greatly across different regions and countries. A model trained on data from one region may not perform well when applied elsewhere, reducing the generalizability of the system.

IV. ALGORITHM OVERVIEW

In the modeling of accident severity, the choice of algorithm depends on several factors such as dataset characteristics, complexity of relationships, and interpretability of results. After experimenting with various algorithms, including Logistic Regression, Decision Tree, XG Boost, AdaBoost, and Extra Tree Classifier, on the given dataset, it’s crucial to evaluate their performance using metrics like accuracy, precision, recall, and F1 score. Among these algorithms, XGBoost stands out with the highest accuracy (0.967) on the training set and a respectable test accuracy of 0.928.

Additionally, it exhibits strong precision (0.964) and a competitive Jaccard score of 0.827, suggesting robust performance in predicting accident severity.

Therefore, based on its high accuracy, precision, and [7] feature importance, XGBoost emerges as a compelling choice for modeling accident severity, offering reliable predictive capabilities for enhancing road safety.

Feature Name	Feature Importance
Number of Vehicles	0.079460
Number of Casualties	0.069389
Day of Week	0.087473
1st Road Class	0.063568
1st Road Number	0.058787
Road Type	0.058984
Speed Limit	0.062032
Junction Detail	0.039503
Junction Control	0.055724
2nd Road Class	0.053979
2nd Road Number	0.014872
Pedestrian Crossing - Human Control	0.001815
Pedestrian Crossing - Physical Facilities	0.022369
Light Conditions	0.061951
Weather Conditions	0.059739
Road Surface Conditions	0.062793
Special Conditions at Site	0.005867
Carriageway Hazards	0.004769
Urban or Rural Area	0.059102
Time Encoded	0.077824

Feature Importance

Model	Train Accuracy	Test Accuracy	Train Precision	Jaccard Score
Logistic Regression	0.800	0.971	0.800	-
Decision Tree	0.800	0.970	0.800	-
XGBoost	0.967	0.928	0.964	0.827
AdaBoost	0.800	0.970	0.800	0.953
Extra Tree Classifier	0.967	0.917	0.970	0.958

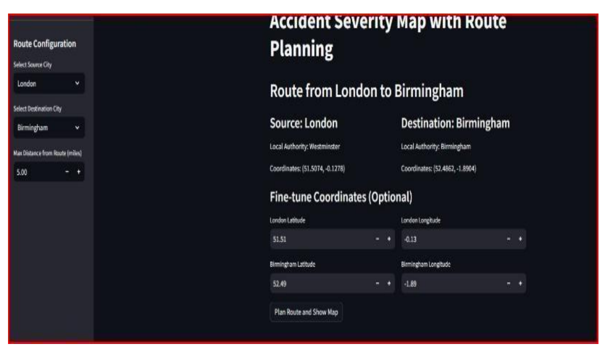
Performance Metrics of Different Models.

XGBoost is a versatile machine learning algorithm widely used for both classification and regression tasks. It operates on the principle of ensemble learning, leveraging the strength of multiple decision trees to enhance predictive performance. XGBoost aggregates predictions from multiple decision trees trained on various subsets of the dataset, improving the overall predictive accuracy by taking the average of the individual predictions. Unlike relying solely on a single decision tree, XGBoost combines the outputs from multiple trees to generate the final prediction. It has demonstrated superior performance, achieving the best results in a significant proportion of applications.

V. KEY FEATURES

This project introduces a comprehensive system designed to enhance road safety by predicting accident severity and suggesting safer travel routes. One of the core features is the severity ranking system, which analyzes historical accident data, traffic patterns, and road conditions to classify accident severity levels on a scale from 1 to 5, where 1 indicates low risk and 5 denotes extremely hazardous zones. By taking user inputs such as the source and destination cities, the system predicts the safest route using machine learning algorithms and provides route optimization that prioritizes lower-risk paths. Additionally, it offers URL generation for route visualization, allowing users to open the suggested path directly in Google Maps for a more interactive and real-time navigation experience. Through this intelligent integration of data analytics, accident prediction, and mapping technologies, the project empowers users to make informed travel decisions and avoid accident-prone areas.

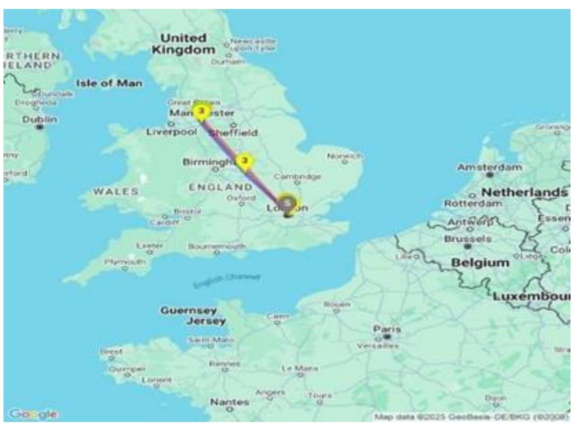
VI. RESULTS OF IMAGE UPLOAD AND ERROR MANAGEMENT



Route Configuration Interface

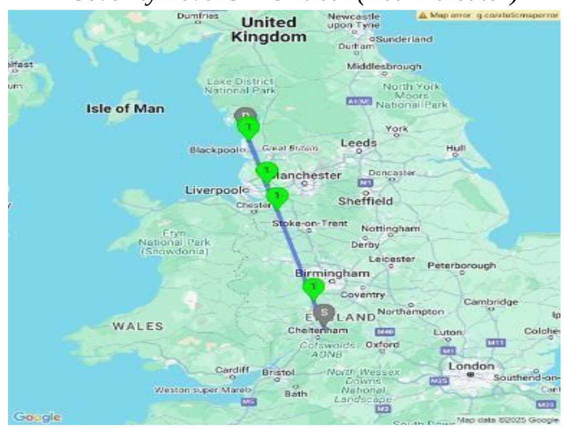
The process of identifying hotspots along the navigation route involves a combination of historical accident data and real-time information. By merging historical and real-time data, the system applies advanced algorithms to identify patterns and trends in accident occurrence along the navigation route. Areas with a higher frequency or severity of accidents are flagged as hotspots. Once identified, the system incorporates this hotspot information into its routing algorithm to provide users with route options that prioritize safety. This may involve suggesting alternate routes that avoid known hotspots and suggests best routes.

Severity zones between source and destination



Severity levels in three alternative routes

Severity Level 5 – Critical (Red Indicator)



Severity Level 1 - Normal (Green Indicator)

VII. CONCLUSION

In summary, the utilization of machine learning algorithms for modeling accident severity in various areas offers a promising solution for improving road safety. By analyzing historical accident data and employing predictive modeling techniques, this approach provides valuable insights into high-risk areas and facilitates targeted interventions. Through the systematic assessment of accident severity, authorities can effectively allocate resources and implement measures to reduce the occurrence and impact of accidents, ultimately saving lives and enhancing road safety on a broader scale.

REFERENCES

[1] Hamzah Al Najada and Imad Mahgoub. Big vehicular traffic data min- ing: Towards accident and congestion prevention. In 2016

International Wireless Communications and Mobile Computing Conference (IWCMC), pages 256–261, 2016.

[2] Meshal Almoshaogeh. Application of XGBOOST model for traffic accidents severity prediction in qassim province of saudi arabia. In 2022 International Conference on Engineering and Emerging Technologies (ICEET), pages 1–6, 2022.

[3] Tadesse Kebede Bahiru, Dheeraj Kumar Singh, and Engdaw Ayalew Tessfaw. Comparative study on data mining classification algorithms for predicting road traffic accident severity. In 2018 Second International Conference on Inventive Communication and Computational Technolo- gies (ICICCT), pages 1655–1660, 2018.

[4] Mateja Iveta, Aleksander Radovan, and Branko Mihaljevic´. Prediction of traffic accidents severity based on machine learning and multiclass clas- sification model. In 2021 44th International Convention on Information, Communication and Electronic Technology,2021.

[5] Kartik Katoch and Raj Gaurang Tiwari. Machine learning-based classification of road accident severity in india: A comparative analysis of algorithms. In 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), pages 1–5, 2023.

[6] Sumbal Malik, Hesham El Sayed, Manzoor Ahmed Khan, and Muham- mad Jalal Khan. Road accident severity prediction — a comparative analysis of machine learning algorithms. In 2021 IEEE Global Confer- ence on Artificial Intelligence and Internet of Things (GCAIoT), pages 69–74, 2021.

[7] Imad El Mallahi, Asmae Dlia, Jamal Riffi, Mohamed Adnane Mahraz, and Hamid Tairi. Prediction of traffic accidents using XGBOOST model. In 2022 International Conference on Intelligent Systems and Computer Vision (ISCV), pages 1–7, 2022.

[8] V Prasad, Parimala Vellivel, and R Chandru. Machine learning algo- rithms to model crash severity and collision(s). In 2022 IEEE North Kar- nataka Subsection Flagship International Conference (NKCon), pages 1–5, 2022.

[9] Xiaoyan Shen and Shanshan Wei. Application of xgboost for haz- ardous material road transport accident severity analysis. IEEE Access, 8:206806–206819, 2020.

[10] Bei Zhou, Xiqing Wang, Shengrui Zhang, Zongzhi Li, Shaofeng Sun, Kun Shu, and Qing Sun. Comparing factors affecting injury severity of passenger car and truck drivers. IEEE Access, 8:153849–153861, 2020.