Vol 7 No 3 (2025) E-ISSN: 2682-860X

International Journal on Robotics, Automation and Sciences

Forecasting High-Risk Traffic Zones Using Machine Learning for Enhanced Road Safety

Nur Farah Nabila Binti Ramzair, Mohammed NasserAl-Andoli* and Cheng Zheng

***Abstract* – Road traffic accidents continue to pose serious global public health and economic challenges. In Malaysia alone, traffic-related incidents caused an estimated RM25 billion in losses in 2023. This study presents a two-part machine learning framework: Part A focuses on predicting accident severity, while Part B uses these predictions to forecast high-risk traffic zones through spatial and temporal analysis. Accident data from 2023 was selected from the UK Road Safety dataset to reflect current traffic patterns, infrastructure, and enforcement efforts. Five classifiers, Logistic Regression, Decision Tree, Random Forest, XGBoost, and K-Nearest Neighbors, were trained and evaluated. A stacking ensemble combining the top three models was constructed to enhance predictive accuracy. The models were assessed using accuracy, precision, recall, and F1-score, with results showing that the ensemble method outperformed individual classifiers. The findings demonstrate the potential of ensemble learning in identifying high-risk zones and supporting proactive road safety planning.**

Keywords— Road Safety, Accident Severity Prediction, Ensemble Learning, Machine Learning, Traffic Risk Mapping, Classification Models.

I. I. INTRODUCTION

Road traffic accidents (RTAs) remain one of the most pressing global health and safety issues. The World Health Organization (WHO) reports that over 1.19 million people die annually, with an additional 20–50 million sustaining injuries of varying severity [1]. Despite improvements in vehicle technologies and infrastructure, RTAs continue to be the eighth leading cause of death worldwide and are projected to rise to the seventh by 2030 without intensified interventions [1,2]. In response, the United Nations proclaimed the Decade of Action for Road Safety 2021–2030, with an ambitious goal of reducing global road traffic deaths and injuries by 50% by 2030 [1,3]. These figures highlight the persistent gap between policy aspirations and the realities of global traffic safety.

In Malaysia, the burden of RTAs remains significant. According to the Malaysian Institute of Road Safety Research (MIROS), the economic loss per road fatality is estimated at RM3.12 million under the Value of Statistical Life (VSL) framework [4,5]. Recent government statistics indicate that road accidents accounted for RM25 billion in losses in 2023, equivalent to approximately 1.4% of the national GDP [6]. With an average of 18 deaths occurring daily, RTAs pose not only a severe public health threat but also a substantial economic burden on the nation's development trajectory.

The persistence of these losses can be attributed to a complex interplay of factors, including driver behavior, infrastructure design, vehicle safety standards, and environmental conditions [7,8]. Traditional statistical methods often struggle to capture the nonlinear and high-dimensional interactions among these variables, limiting their ability to generate reliable predictive insights. Recent studies demonstrate that machine learning (ML) approaches offer a more robust alternative, with the capacity to model complex interdependencies and uncover latent patterns in traffic data, thereby enabling proactive and adaptive safety interventions [8–10]. The integration of ML into traffic risk modeling thus represents a promising avenue for advancing both predictive accuracy and policy relevance in road safety research.

Corresponding Author email: nasser.alandoli@mmu.edu.my, ORCID: 0000-0001-6491-9938
Nur Farah Nabila is with Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia (e-mail: 1221301140@student.mmu.edu.my).
Mohammed Nasser is with Centre for Cybersecurity and Quantum Computing, CoE for Advanced Cloud, Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia, Multimedia University, Cyberjaya, Malaysia (e-mail: nasser.alandoli@mmu.edu.my).
Cheng Zheng is with Wireless Signal Processing Technology Inc, Canada. (e-mail: zheng.cheng@wirelessignal.com)

International Journal on Robotics, Automation and Sciences (2025) 7, 3:49-57
https://doi.org/10.33093/ijoras.2025.7.3.7
Manuscript received: 3 Aug 2025 | Revised: 21 Sep 2025 | Accepted: 28 Sep 2025 | Published: 30 Nov 2025

Published by MMU PRESS. URL: http://journals.mmupress.com/ijoras

This study proposes a machine learning-based framework for forecasting high-risk traffic zones using the 2023 UK Road Safety dataset. The dataset was deliberately scoped to a single year to ensure alignment with current infrastructure, enforcement trends, and behavioral shifts following the COVID-19 pandemic. As noted by Reddy et al. (2024), the use of recent data enhances relevance and decision-making accuracy in real-time traffic safety analytics [4]. However, the dataset exhibits significant class imbalance, with most accidents classified as Slight and relatively few classified as Fatal or Serious. This imbalance presents a technical challenge for multi-class classification models aiming to deliver balanced predictive performance across severity levels.

To address these limitations, the study is structured into two phases. Part A focuses on accident severity prediction, where five classifiers including Logistic Regression, Decision Tree, Random Forest, XGBoost, and K-Nearest Neighbors, are trained and evaluated. A stacking ensemble model is then constructed by combining the top-performing classifiers to enhance overall accuracy and sensitivity to minority classes. Part B extends the analysis through spatial and temporal forecasting, using severity probabilities to identify high-risk zones via heatmaps and grid-based mapping. The contributions of this paper are threefold: (1) a comparative analysis of five machine learning models using a single-year, real-world accident dataset; (2) the development of a stacking ensemble to address class imbalance and improve predictive performance; and (3) the application of geospatial techniques to forecast high-risk areas and support data-driven road safety planning.

II. RELATED WORKS

Road traffic accident severity prediction has seen a marked shift from traditional statistical models to more powerful machine learning (ML) approaches. Early methods like logistic regression provided interpretability but struggled with complex, high-dimensional datasets.

A prominent improvement has been achieved using tree-based models such as Random Forest (RF) and XGBoost, which handle nonlinearity and feature interactions effectively. For example, a study in Saudi Arabia demonstrated that XGBoost outperformed RF and logistic regression using accident data from Qassim Province, achieving up to 94% accuracy and an AUC of 0.98 in binary classification [11]. Another balanced and unbalanced dataset study reported that stacking multiple base models with logistic regression as the meta-learner achieved an AUC of 96.92%, outperforming individual classifiers [12].

Ensemble learning, especially stacking, has been explored extensively. A two-layer stacking model combining RF, AdaBoost, and GBDT (Gradient Boosting Decision Trees) effectively predicted crash injury severity [13]. Recent efforts also investigated stacking strategies in ensemble frameworks for crash frequency forecasting, demonstrating superior out-of-sample predictive performance over parametric models [14]. Additional work reinforced the robustness of two-layer stacking models in predicting traffic accident severity [15].

Some studies compared tree-based and ensemble regression approaches for severity prediction. One such analysis revealed that Random Forest achieved exceptional results, with accuracy of 0.974, precision 0.954, recall 0.930, and F1 score 0.942, outperforming other models like AdaBoost and Gradient Boosting [7].

III. METHODOLOGY

This study adopts a two-phase methodology aimed at predicting traffic accident severity and identifying high-risk spatial zones using machine learning techniques. Phase A focuses on developing supervised classification models for accident severity prediction, while Phase B leverages model probability outputs to generate severity-based heatmaps and spatial risk maps. The complete workflow is illustrated in Fig. 1.

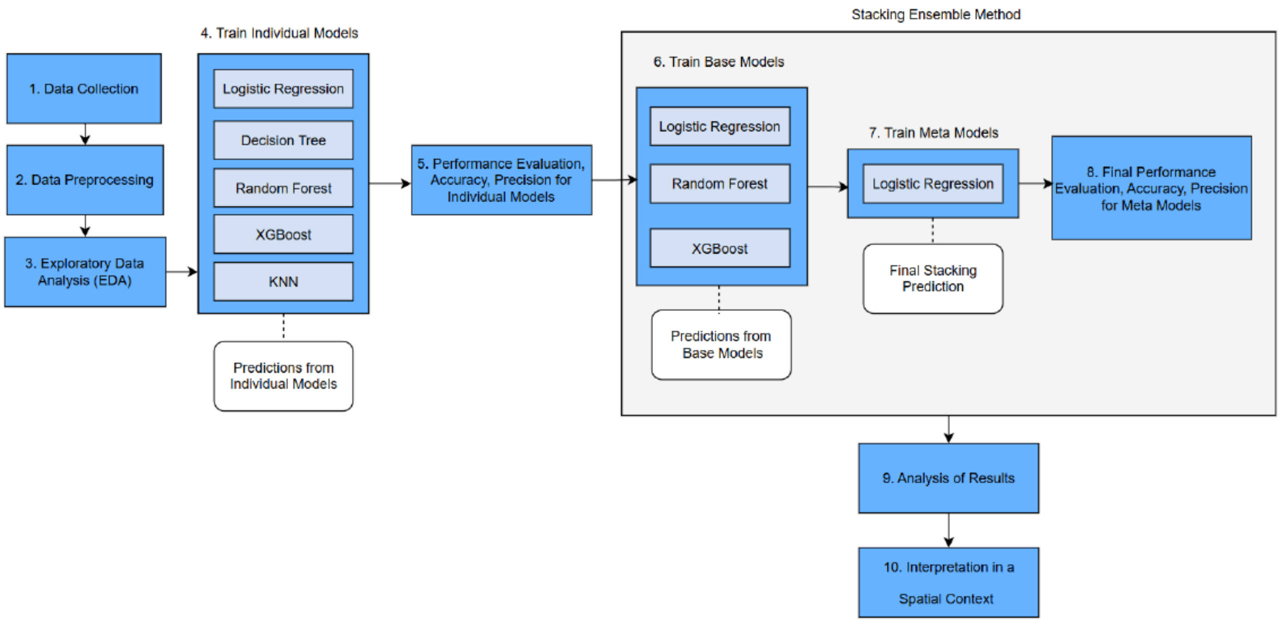


Figure. 1. Workflow of the proposed methodology for accident severity prediction and spatial risk mapping.

1) Dataset

The Road Safety Data – Collisions 2023 dataset, published by the UK Department for Transport [9], was utilized. It contains 104,258 records with 37 attributes, covering spatial, temporal, environmental, and accident-related features. After preprocessing, the dataset was refined to 19,926 clean records and 22 predictive attributes (Table I). Key features include accident location (longitude/latitude), number of vehicles involved, timestamp, weather conditions, light conditions, and road surface type.

Table 1. Attribute description.

No.	Attribute	Description
1	Longitude	Longitude of the location of an accident scene.
2	Latitude	Latitude of the location of an accident scene.
3	Vehicles	Number of vehicles involved in the accident
4	Date	Date of the accident.
5	Day of Week	Day of the week that accident occurred.
6	Time	Timestamp of the accident.
7	Weather	Weather condition at the time of the accident.
8	Light Condition	Light conditions at the time of the accident.
9	Road Surface	Road Surface at the time of the accident.
10	Urban/ Rural Area	Area where the accident occurred.

2) Data Preprocessing

To ensure data quality and improve model performance, several preprocessing steps were performed:

a) Missing and Duplicate Values: Initial inspection revealed several columns with missing entries, including *accident index, accident reference*, and *date*. In total, 3,303 rows with missing critical identifiers and 61,122 rows missing date information were removed, reducing the dataset from 104,258 to 37,863 instances. Additionally, duplicate entries were removed to maintain data integrity.

b) Feature Selection**:** From the original 37 features, non-predictive or redundant attributes such as *accident_index*, *accident_reference*, and coordinate references were excluded. After feature selection, the dataset was reduced to 22 relevant attributes, retaining only those with predictive or spatial significance.

c) Outlier Removal**:** Outliers in selected numerical features such as *number_of_vehicles, speed_limit,* and *total_trans_amt*, were detected using the Interquartile Range (IQR) method. The IQR is defined in Eq. (1)

$$IQR = Q3 - Q1 \quad (1)$$

Acceptable values were bound by:

$$[Q1 - 1.5 \times IQR, Q3 + 1.5 \times IQR] \quad (2)$$

This filtering removed 19,907 extreme rows, reducing the dataset from 39,833 to 19,926 clean records.

d) Feature Scaling: To standardize the scale of numeric variables and improve model performance, Min- Max normalization was applied. Each value was transformed using Eq. (3):

$$X_{scaled} = X - X_{min} / X_{max} - X_{min} \quad (3)$$

This resolution ensures that all features contribute equally during model training. The original *speed limit* values were preserved in a separate column for interpretability during post-analysis.

3) Exploratory Data Analysis

Exploratory Data Analysis (EDA) was conducted to gain insights into the distribution, frequency, and characteristics of traffic accidents in the dataset. The primary focus was on understanding accident severity distribution, temporal patterns, environmental factors, and spatial trends.

a) Severity Distribution: The dataset exhibits a severe class imbalance, with the majority of cases categorized under *Slight* severity as. Fig.2 showed *Serious* and *Fatal* cases occur far less frequently, which aligns with real-world accident statistics. This imbalance poses a significant challenge for classification models, particularly in accurately identifying high-risk events.

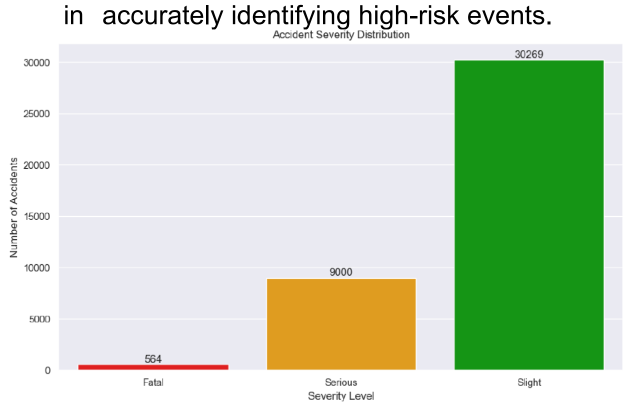


Figure. 2. Accident severity distribution.

b) Temporal Trends**:** Accidents were analyzed across months, days of the week, and time intervals. Fig 3. Illustrated that higher accident frequencies were observed during weekdays, particularly during morning and evening rush hours, suggesting a strong link to commuting activity. Month-wise trends revealed notable seasonal fluctuations, which may reflect varying traffic volumes or weather conditions.

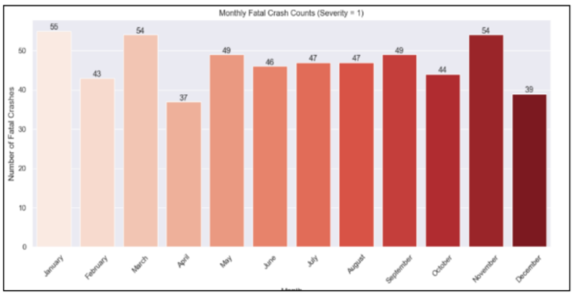


Figure 3. Monthly distribution of fatal traffic crashes (Severity = 1).

c) Environmental Conditions: Variables such as *weather_conditions, light_conditions*, and *road_surface_conditions* were examined for correlations with accident severity. A higher proportion of *Serious* and *Fatal* accidents occurred under poor lighting or wet road conditions, indicating the influence of environmental risk factors on accident outcomes.

d) Geospatial Patterns: Using the *longitude* and *latitude* attributes, accident locations were visualized to identify spatial clusters. Dense accident hotspots were predominantly located in urban centers, consistent with areas of high traffic flow and population density. These visualizations supported later development of severity heatmaps and grid-based spatial risk maps.

4) Individual Models Training

To establish robust predictive baselines, five widely used supervised machine learning classifiers were implemented: Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Extreme Gradient Boosting (XGB), and K-Nearest Neighbors (KNN). These models were chosen to capture a broad spectrum of learning paradigms, ranging from linear (LR) to non-linear tree-based (DT, RF, XGB) and instance-based learning (KNN).

The dataset was partitioned into training and testing subsets using an 80/20 stratified split, ensuring the preservation of the original class distribution. To mitigate the impact of class imbalance, model training pipelines incorporated a ColumnTransformer for numerical feature scaling and categorical one-hot encoding, followed by a hybrid resampling strategy combining Synthetic Minority Over-sampling Technique (SMOTE) with Random Under-Sampling (RUS). This ensured an enhanced representation of minority classes without excessive duplication of synthetic samples.

Hyperparameter optimization was conducted using both GridSearchCV and RandomizedSearchCV, applying 3–5 fold cross-validation depending on model complexity. This dual strategy balanced exhaustive search with computational efficiency. Label encoding was selectively applied for algorithms requiring integer-encoded labels (e.g., XGBoost)..

5) Stacking Model Training

Building on the performance of individual models, a stacking ensemble architecture was designed to exploit model diversity and enhance generalization. Logistic Regression, Random Forest, and XGBoost were selected as base learners due to their complementary strengths in linear discrimination, bagging-based variance reduction, and boosting-based bias correction.

The meta-learner was implemented using Gradient Boosting, trained on the class probability outputs of the base models. A two-stage training approach was employed: base learners were trained on stratified folds, and the meta-learner was trained exclusively on out-of-fold predictions, thereby preventing data leakage and ensuring an unbiased estimation of ensemble performance.

This hierarchical architecture enables the ensemble to capture heterogeneous decision boundaries across classifiers, ultimately improving robustness against class imbalance and enhancing sensitivity to rare but critical accident severity classes.

6) Spatial Analysis

a) Heatmap Visualization: To identify localized accident risk, severity probabilities derived from the predictive models were spatially mapped using the Folium heatmap library. Each accident record was weighted by its predicted severity probability, thereby emphasizing regions with higher concentrations of severe accident likelihoods. This probabilistic weighting allowed the visualization to highlight urban hotspots where severe accidents are more prevalent, aligning spatial patterns with population density and traffic congestion levels.

b) Grid-Based Aggregation: To facilitate regional-level analysis and mitigate the noise inherent in individual accident locations, a grid-based spatial aggregation approach was applied. Geographic coordinates (latitude, longitude) were rounded to 0.01° precision, corresponding to approximately 1.1 km × 1.1 km cells. For each cell, the mean severity probability was computed to represent localized accident risk intensity. These aggregated scores were visualized as a choropleth risk map using a sequential YlOrRd color gradient, where darker shades denoted higher predicted risk. This transformation enabled systematic hotspot detection and comparative assessment across urban and rural zones.

c) Composite Risk Score: To capture both the frequency of accidents and their predicted severity, a composite spatial risk score was formulated (Eq. 4).

$$Composite\ Score = 0.7 \times Avg\ Severity + 0.3 \times (A\frac{ccident\ Count}{Max\ Count}) \quad (4)$$

This dual-factor measure integrates the number of incidents within a grid cell and their associated severity probabilities, providing a holistic indicator of regional accident risk. The composite index not only highlights areas with frequent accidents but also prioritizes regions where the potential consequences are more severe, thereby offering actionable insights for resource allocation, traffic enforcement, and urban planning interventions

7) Model Evaluation

Model performance was assessed using four key metrics: Accuracy**,** Precision, Recall, F1 Score and Confusion Matrix. These metrics offer a balanced view of overall performance and the model's sensitivity to critical cases. Accuracy measures the proportion of correctly predicted samples, Precision, and Recall, particularly for the Fatal class, highlights how well the model detects severe incidents. F1-Score balances precision and recall. Lastly, The Confusion Matrix was used to visualize the classification results and reveal misclassifications across severity levels.

On severe outcomes, a composite severity probability was calculated using Eq. (5) by summing the probabilities of the Fatal and Serious classes:

$$Predicted\ Severity\ Probability = P(Fatal) + P(Serious) \quad (5)$$

Accidents with a predicted severity probability greater than or equal to 0.3391 were flagged as high-risk. This threshold was determined based on model calibration and adjusted to balance sensitivity and specificity. Two new fields, *predicted_severity_prob* and a binary *high_risk* indicator, were added to the dataset for subsequent analysis.The average severity probability across the dataset was also computed as a baseline indicator of overall risk is define in Eq. (6):

$$avg = \frac{P_{fatal,1} + P_{serious,1} + \cdots + P_{fatal,n} + P_{serious,n}}{n} \quad (6)$$

IV. RESULTS AND DISCUSSION

1) Spatial Distribution of Predicted Accident Severity

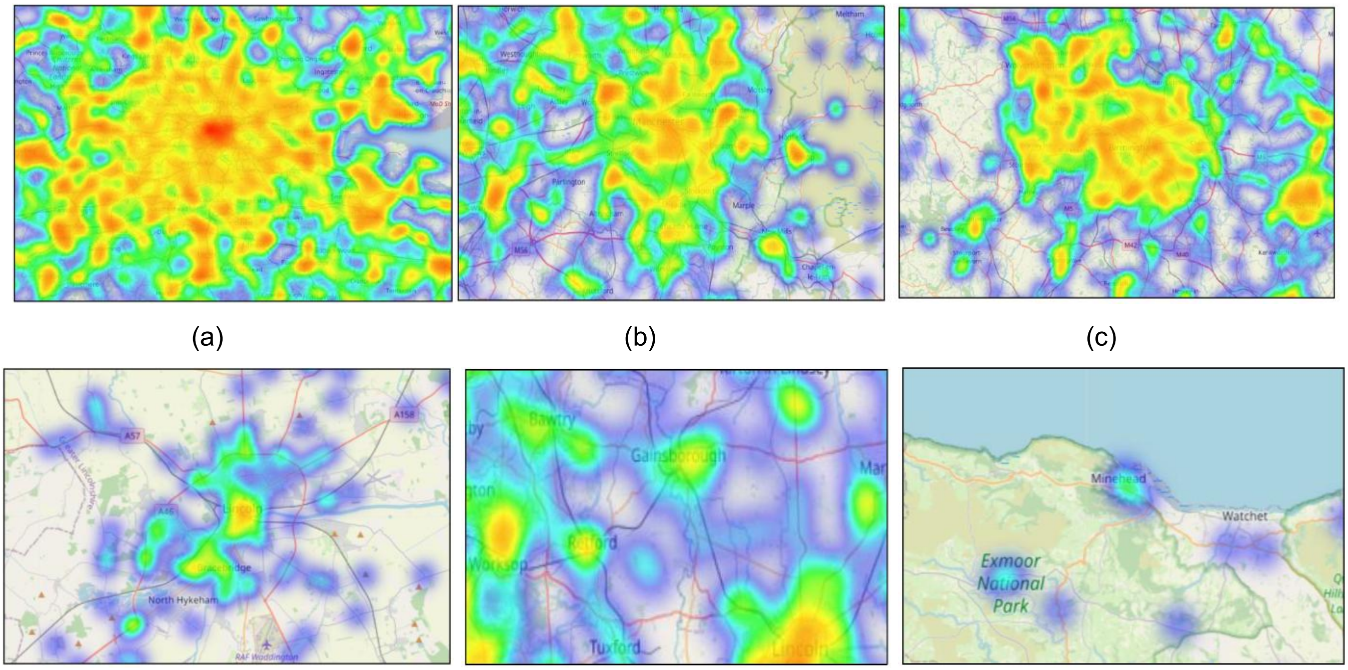
The spatial analysis of predicted accident severity revealed distinct geographic patterns when visualized using severity-weighted heatmaps. In Figure 4a, London stands out as the most intense hotspot, characterized by deep red clusters that highlight consistently high predicted severity. This concentration is likely driven by the city's dense traffic networks, high vehicle volumes, and complex road structures that amplify the risk of serious collisions.

Beyond the capital, additional clusters emerged in regional urban centers such as Manchester and Birmingham (Figures 4b–4c). These zones, represented by orange to yellow gradients, indicate moderate-to-high severity levels. The patterns in these cities suggest a strong influence of high-speed routes and regional intersections that increase the likelihood of severe outcomes.

At a finer spatial scale, localized "silent risk zones" were detected in smaller towns such as Lincoln and Gainsborough (Figures 4d–4e). Although these areas report fewer accidents in terms of frequency, the model assigns them elevated severity probabilities. This finding underscores the value of probabilistic mapping, as such regions might otherwise remain overlooked in traditional frequency-based analyses.

By contrast, rural and less densely populated areas such as Watchet and Witham displayed predominantly blue and green zones, signifying lower predicted severity. These patterns likely reflect reduced traffic intensity and simpler road structures, although the limited availability of training data in rural regions may also affect the model's predictive confidence.

Overall, the combined spatial heatmap demonstrates that severity-based risk assessment can effectively highlight both established urban hotspots and emerging high-risk pockets. This geospatial perspective provides actionable insights for policymakers, supporting targeted interventions such as infrastructure upgrades, dynamic traffic regulations, and localized road safety campaigns.



(a) (b) (c)

(d) (e) (f)

Figure 4. Spatial distribution of predicted accident severity across the UK.

(a) London, showing the most intense severity hotspots;
(b) Manchester, highlighting moderate-to-high severity clusters;
(c) Birmingham, with regionally concentrated risk zones;
(d) Lincoln, and (e) Gainsborough, illustrating localized "silent risk zones" with elevated severity despite lower accident frequency;
(f) Severity is visualized using a probabilistic heatmap, where red indicates high predicted severity (Fatal + Serious), while blue-green represents lower predicted severity.

2) *Grid-Based Aggregation for Localized Risk Detection*

To uncover structured spatial patterns beyond individual accident points, predicted severity probabilities were aggregated into uniform 0.01° grid cells (≈1.1 km²). Each cell's average severity was mapped using a sequential yellow-to-red colormap, transforming scattered events into a coherent spatial framework. This grid-based approach highlights both concentrated and diffuse high-risk zones, with deep-red clusters around urban Birmingham, transitional orange/yellow cells in suburban corridors, and lighter shades across rural areas. The resulting visualization (Figure 5.27) provides a clearer basis for identifying persistent accident hotspots and prioritizing localized safety interventions.

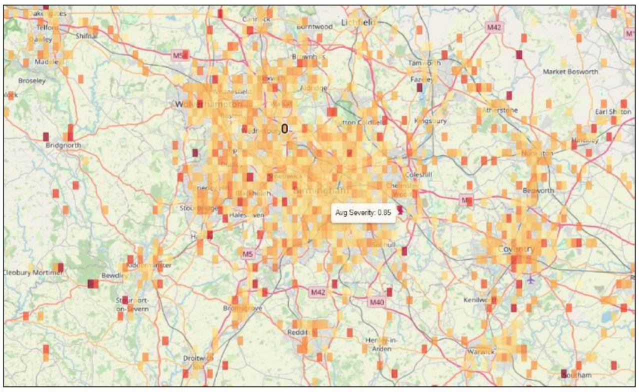


Figure 5. Grid map of urban Birmingham surrounded by red, orange and yellow rectangles.

Beyond enhancing visual clarity, the grid-based framework increases actionability by quantifying accident risk at a micro-geographic scale. This makes the approach particularly valuable for policymakers, urban planners, insurers, and smart city stakeholders, who can use the results to prioritize high-severity zones for targeted interventions. Practical applications include deploying traffic-calming measures in red-zone cells, prioritizing infrastructure upgrades in high-risk areas, dynamically adjusting insurance premiums, and allocating emergency response resources more efficiently.

	lat_bin	lon_bin	accident_count	avg_severity	composite_score
11513	52.60	-1.71	1	0.994158	0.703053
10238	52.36	-2.37	1	0.994023	0.702959
5723	51.50	-1.87	1	0.991989	0.701535
133	50.31	-5.03	1	0.991438	0.701149
3422	51.26	1.28	1	0.991071	0.700892
4768	51.42	-0.56	1	0.991071	0.700892
7356	51.65	-2.65	1	0.990463	0.700467
11648	52.62	-0.15	1	0.989370	0.699702
11591	52.61	-0.20	1	0.989357	0.699692
11249	52.55	1.67	1	0.989264	0.699628

Figure 6: Composite risk score.

A key innovation is the introduction of a composite risk score (Figure 6), which integrates both severity probability (70%) and accident frequency (30%). This weighting ensures that even locations with rare but potentially catastrophic incidents are flagged as critical, preventing reliance on frequency alone. The top 10 highest-risk cells, all with severity probabilities near 1.0 despite low accident counts, underscore the model's sensitivity to silent but high-impact risks.

Together, the grid-based maps and composite scoring system provide a dual-layered decision support tool: the maps highlight broader spatial risk patterns, while the scores pinpoint micro-locations requiring immediate attention. This duality balances macro-level insight with micro-level prioritization, offering a practical pathway toward proactive, data-driven road safety strategies. Moreover, because the method does not depend on fixed administrative boundaries, it remains scalable across different jurisdictions and adaptable to diverse urban and regional contexts.

3) *Risk Trend for Top 3 Zones*

To complement spatial risk mapping, this section examines the temporal evolution of accident severity across grid zones. Severity probabilities predicted by the stacking ensemble were aggregated monthly and annually at a 0.1° grid resolution, enabling structured tracking of risk fluctuations over time. A pivot table was used to visualize monthly severity trends, and the three highest-risk grids (by average severity) were selected for deeper analysis.

The resulting trends (Figure 7) reveal that accident severity is strongly influenced by seasonal factors rather than being randomly distributed. Peaks are observed in winter months (December–January), likely due to adverse weather and reduced daylight, while additional spikes in summer (July–August) suggest links to increased travel, tourism, and congestion. Importantly, each zone displays distinct temporal dynamics, highlighting the location-specific nature of risk patterns.

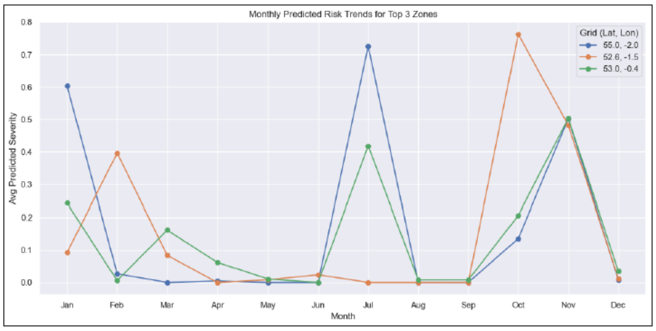


Figure 7: Line chart of monthly predicted risk trends for Top 3 zones.

Zone 1 (Lat: 55.0, Lon: –2.0) shows a volatile profile, with high severity in January (0.6), a decline through spring, and a sharp July spike above 0.7, alongside secondary peaks in October–November. Zone 2 (Lat: 52.6, Lon: –1.5) demonstrates an autumn-centric pattern, with notable surges in October (>0.7) and November (~0.5). Zone 3 (Lat: 53.0, Lon: –0.4) remains relatively stable but exhibits a modest November increase (~0.5). Cross-zone comparisons show a consistent rise in November severity, suggesting system-wide seasonal vulnerability, whereas September consistently marks a low-risk period.

From an operational standpoint, these results underscore the value of time-aware risk management. Seasonal variation implies that safety interventions should be scheduled dynamically: for example, summer-focused measures in Zone 1, autumnfocused strategies in Zone 2, and targeted year-end preparedness in Zone 3. The identification of a universal November risk peak highlights the need for coordinated national responses, while the September lull may represent a window for resource reallocation.

4) *Performance Analysis of Individual and Ensemble Models*

The results in Table 2 highlight clear trade-offs between the evaluated classifiers. XGBoost (XGB), an advanced ensemble learning algorithm based on gradient boosting, achieves the highest accuracy (0.73) and demonstrates strong recall for the majority "Slight" class (0.91). However, its performance drops significantly for minority classes, with recall values of only 0.07 for Fatal and 0.15 for Serious accidents. This suggests that while boosting effectively optimizes performance on the dominant class, it struggles to generalize to underrepresented but critical categories.

RF, another ensemble-based method using bagging, provides more balanced predictions, achieving a macro-average F1 of 0.40, though with lower overall accuracy compared to XGB. LR and KNN contribute complementary strengths, with LR achieving relatively high Fatal recall (0.60) and KNN performing better for Serious recall (0.41), though both models are limited in overall accuracy and consistency.

The Stacking Ensemble (SEL) combines the strengths of LR, RF, and XGB under a meta-classifier, producing a more equitable trade-off between accuracy and class-specific recall. While its accuracy (0.71) is slightly lower than XGB, it achieves markedly higher recall for Fatal (0.37) and Serious (0.41) classes. This demonstrates the value of stacking in leveraging heterogeneous learners: whereas XGB alone tends to favor the majority class, SEL ensures broader generalization by integrating multiple decision boundaries.

In summary, although XGB, as an ensemble learning approach that achieves the strongest overall accuracy, the stacking ensemble (SEL) delivers a superior balance by capturing rare but high-severity accidents more effectively. In the context of road safety, this balanced predictive capability is more valuable than raw accuracy, since Fatal and Serious cases, despite their rarity, are the most critical for intervention and policymaking.

Table 2. Performance comparison of supervised classifiers and stacking ensemble (SEL) for accident severity prediction.

Model	Accu.	Weighted Avg F1	Class 1 (Fatal) Recall	Class 2 (Serious) Recall	Class 3 (Slight) Recall	MacrAvg F1
LR	0.49	0.56	0.6	0.3	0.54	0.33
DT	0.61	0.64	0.3	0.31	0.71	0.38
RF	0.64	0.66	0.34	0.32	0.74	0.4
XGB	0.73	0.69	0.07	0.15	0.91	0.37
KNN	0.55	0.59	0.15	0.41	0.6	0.35
SEL	**0.71**	**0.68**	**0.37**	**0.41**	**0.90**	**0.38**

The confusion matrix (Figure 8) confirms these findings. It shows that while the model still tends to misclassify severe accidents as "Slight," this behavior is now more predictable and consistent, rather than erratic. Fatal and Serious misclassifications largely occur within adjacent severity levels, suggesting that the model has learned to distinguish broad severity tiers, even if it struggles with fine-grained classification.

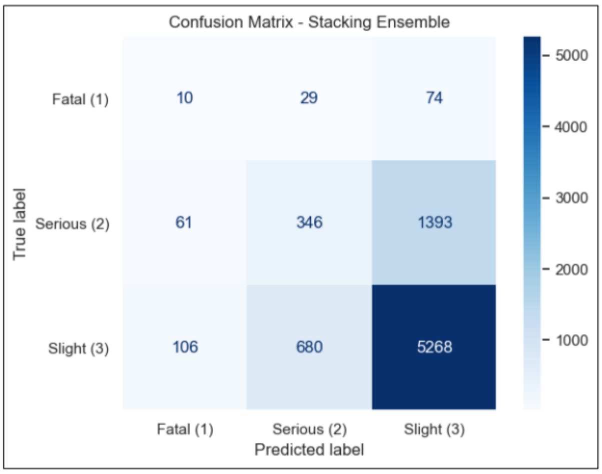


Figure 8. Confusion matrix of Stacking Ensemble Model.

5) *Discussion*

The study demonstrates that machine learning, particularly ensemble approaches, can substantially enhance the prediction and interpretation of road accident severity. XGBoost achieved the highest accuracy, confirming the strength of boosting methods in optimizing dominant patterns, while the stacking ensemble offered a more balanced performance by improving recall for Fatal and Serious cases—two categories that are critical for policymaking despite their rarity. This highlights a key achievement: the framework does not simply maximize accuracy but enhances equity in risk detection across severity levels, thereby increasing its real-world utility.

Spatially, severity-weighted heatmaps and grid-based aggregation revealed both urban hotspots and localized "silent risk zones," providing a dual-layered system for macro- and micro-level safety planning. The introduction of a composite risk score further refined hotspot detection by combining severity probability with frequency, ensuring that rare but catastrophic risks were not overlooked. Temporally, the analysis uncovered seasonal trends, with severity peaking in winter and summer and showing a recurring national rise in November, underscoring the importance of time-aware risk mitigation strategies.

Nevertheless, limitations remain. Class imbalance hindered minority recall, particularly for XGBoost, while sparsity in rural data reduced prediction confidence in low-density areas. Additionally, the models relied solely on collision data, excluding potentially influential variables such as weather, driver demographics, and vehicle characteristics.

Future research should integrate multimodal datasets, employ explainable AI for interpretability, and develop real-time predictive systems that adapt dynamically to evolving traffic conditions. These advancements will not only improve prediction accuracy but also foster actionable, trustworthy insights for intelligent transport systems and policy interventions. Final, we plan to extend the evaluation to include temporal and spatial generalization tests and a detailed analysis of model training and inference times to better reflect real-world deployment scenarios.

V. CONCLUSION

This study presented a comprehensive framework for predicting and analyzing road traffic accident severity in the UK using supervised machine learning and ensemble methods, integrating both spatial and temporal perspectives. Among the models tested, XGBoost achieved the highest overall accuracy, yet its weakness in recalling minority classes highlighted the limitations of focusing solely on dominant patterns. The stacking ensemble addressed this issue by achieving a more balanced trade-off, substantially improving the detection of Fatal and Serious cases, which, despite their rarity, are of greatest importance to road safety policy. Beyond classification performance, the research advanced accident risk analysis through severity-weighted heatmaps, grid-based aggregation, and the introduction of a composite risk score, which captured both frequency and severity to identify silent high-risk zones that frequency-based methods often overlook. Temporal analysis further revealed that accident severity follows seasonal trends, with recurring high-risk peaks in winter and summer, as well as a nationally consistent spike in November, underscoring the need for time-sensitive interventions. Collectively, these findings demonstrate that road accident severity is not randomly distributed but shaped by spatial, temporal, and contextual factors, and that data-driven frameworks can provide actionable insights for targeted infrastructure upgrades, dynamic policy design, and resource allocation. While challenges remain in addressing class imbalance, rural data sparsity, and integration of external variables such as weather and driver behavior, future research can build on this work by incorporating multimodal data, interpretable AI, and real-time analytics to support intelligent transport systems and smart city risk governance.

ACKNOWLEDGMENT

We thank the reviewers and editor of the journal for their guidance in improving the quality of our article.

FUNDING STATEMENT

There are no funding agencies supporting the research work.

AUTHOR CONTRIBUTIONS

Nur Farah Nabila: Conceptualization, Data Curation, Methodology, Validation, Writing –Original Draft Preparation;

Mohammed Al-Andoli: Project Administration, Supervision , Writing – Review & Editing;

Cheng Zheng: Writing – Review & Editing.

CONFLICT OF INTERESTS

No conflict of interests were disclosed.

ETHICS STATEMENTS

This research did not involve human participants, animal subjects, or sensitive personal data, and therefore did not require ethical approval.

REFERENCES

[1] WHO, “Road Traffic Injuries Fact Sheet; World Health Organization: Geneva,” *World Health Organization*, 2023. URL:https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries (Accessed: 21 August 2025).

[2] M.Å. Belin, M. Khayesi and N. Tran, “‘Road safety is no accident’: building efficient road safety lead agencies, strategies and targets in the world, 2009–2023,” *Injury Prevention*, 2025. DOI: https://doi.org/10.1136/ip-2024-045601.

[3] WHO, “Decade of Action for Road Safety 2021–2030,” *World Health Organization*, 2021. URL:https://www.who.int/teams/social-determinants-of-health/safety-and-mobility/decade-of-action-for-road-safety-2021-2030. (Accessed: 21 August 2025).

[4] J.S. Reddy, M.R. Rashmi and H.P. Lee, “Road accident prediction in highways using machine learning algorithms,” *2024 5th International Conference on Smart Electronics and Communication (ICOSEC)*, pp. 1-6, 2024. DOI: https://doi.org/10.1109/ICOSEC61587.2024.10722164

[5] M.S.A. Samad et al, "Investigation on the effect of Malaysian anthropometric size in vehicle crash safety by using finite element method," *Journal of the Society of Automotive Engineers Malaysia*, vol. 5, no. 3, pp. 449-466, 2021. DOI: https://doi.org/10.56381/jsaem.v5i3.187

[6] A. Loke, “Road Accidents Cost Malaysia RM25 Billion in 2023,” *The Star*, 11 July 2024. URL:https://www.thestar.com.my/news/nation/2024/07/11/road-accidents-cost-malaysia-rm25bil-in-economic-value-in-2023-says-transport-minister (Accessed: 21 August 2025)

[7] M. Sohail, M. Khan and S. Shah, “Behavioral and environmental determinants of road accidents: Insights for safety policy,” *Journal of Safety Research*, vol. 86, pp. 31–41, 2023. DOI: https://doi.org/10.1016/j.jsr.2023.01.004

[8] X. Zhao, Y. Xu and R. Liang, “Traffic accident severity prediction using ensemble machine learning methods,” *Applied Intelligence*, vol. 52, pp. 14478-14495, 2022. DOI: https://doi.org/10.1007/s10489-022-03117-y

[9] Y. Li, H. Wang and J. Sun, “Machine learning applications in traffic accident prediction: A comprehensive review,” *Accident Analysis & Prevention*, vol. 181, p. 106950, 2023. DOI: https://doi.org/10.1016/j.aap.2022.106950

[10] L, Zhang, Q, Chen and H. Zhao, “Deep learning for traffic safety: A review of methods and applications,” *IEEE Transactions on Intelligent Transportation Systems*, 25, 1203–1217, 2024. DOI: https://doi.org/10.1109/TITS.2023.3299847

[11] Aldhari, M. Almoshaogeh, A. Jamal, F. Alharbi, M. Alinizzi and H. Haider, "Severity prediction of highway crashes in Saudi Arabia using machine learning techniques," *Applied Sciences*, vol. 13, no. 1, p. 233, 2022. DOI: https://doi.org/10.3390/app13010233

[12] A. Çelik and O. Sevli, "Predicting traffic accident severity using machine learning techniques," *Türk Doğa ve Fen Dergisi*, vol. 11, no. 3, pp. 79-83, 2022. DOI: https://doi.org/10.46810/tdfd.1136432

[13] J. Tang, J. Liang, C. Han, Z. Li and H. Huang, "Crash injury severity analysis using a two-layer Stacking framework," *Accident Analysis & Prevention*, vol. 122, pp. 226-238, 2019. DOI: https://doi.org/10.1016/j.aap.2018.10.016
[14] N. Ahmad, B. Wali and A.J. Khattak, "Heterogeneous ensemble learning for enhanced crash forecasts–a frequentist and machine learning based stacking framework," *Journal of safety research*, vol. 84, pp. 418-434, 2023.
DOI: https://doi.org/10.1016/j.jsr.2022.12.005
[15] I. Aldhari, M. Almoshaogeh, A. Jamal, F. Alharbi, M. Alinizzi, and H. Haider, "Severity prediction of highway crashes in Saudi Arabia using machine learning techniques," *Applied Sciences*, vol. 13, no. 1, p. 233, 2022.
DOI: https://doi.org/10.3390/app13010233
[16] J.S.V. Leong and K.B. Gan, "Cuffless Non-invasive Blood Pressure Measurement Using CNN-LSTM Model: A Correlation Study," *International Journal on Robotics, Automation and Sciences*, vol. 5, no. 2, pp. 25-32, 2023. DOI: https://doi.org/10.33093/ijoras.2023.5.2.3
[17] T. Tai, S. Haw, W. Kong and K. Ng, "Performance Evaluation of Machine Learning Techniques on Resolution Time Prediction in Helpdesk Support System," *International Journal on Robotics, Automation and Sciences*, vol. 6, no. 2, pp. 59-68, 2024.
DOI: https://doi.org/10.33093/ijoras.2024.6.2.9

.