A Graph Based Hybrid Approach to Injury Severity Prediction in Road Accidents Using Deep Learning

Jyoti B. Bhosale and Bhushan S. Yelure

ABSTRACT

Road traffic accidents are still one of the major problems all over the world that lead to a very high number of injuries and deaths. The accurate measurement of the severity of the injuries caused by these accidents is very essential indeed as a way of upgrading road safety, making the emergency response more efficient, and also, just to name a few, giving policy decisions that are based on evidence. Traditional statistical models, to a great extent, fail to capture all the complex nature of the relationships among different crash-related factors. In order to deal with the GNNRF (Graph Neural Network-Random Forest) is a combined machine learning model introduced by this research work. This model constitutes one of the best hypotheses of deep learning technologies for graphs combined with the strong points of ensemble learning to make predictions on crash injury severity. Essentially, by change crash data sets into graph format through k-nearest neighbor (kNN) connections and PCA-derived embeddings, the model becomes competent in finding hidden interdependencies among crash records. The experimental appreciations provide evidence that the performance of GNNRF is better than that of the traditional methods such as Random Forest and XGBoost, especially when it coherence with accuracy, and other evaluation metrics, such as precision, recall, and F1-score. Further, Explainable AI (XAI) techniques

Keywords: Road crash severity prediction, Graphsage-based GNN, KNN graph with PCA embeddings, GNN-random forest hybrid, Shap explainable AI

Government College of Engineering, Karad, Maharashtra, India ROR

Correspondence to:
Jyoti B. Bhosale,
jyotisalunkhe16@gmail.com

Additional material is published online only. To view please visit the journal online.

Cite this as: Bhosale JB and Yelure BS. A Graph Based Hybrid Approach to Injury Severity Prediction in Road Accidents Using Deep Learning. Premier Journal of Science 2025;14:100131

DOI: https://doi.org/10.70389/PJS.100131

Peer Review

Received: 19 August 2025

Last revised: 24 September 2025

Accepted: 29 September 2025

Version accepted: 6

Published: 27 December 2025

Ethical approval: N/a

Consent: N/a

Funding: No industry funding

Conflicts of interest: N/a

Author contribution: Jyoti B. Bhosale and Bhushan S. Yelure – Conceptualization, Writing – original draft, review and editing

Guarantor: Jyoti B. Bhosale

Provenance and peer-review: Unsolicited and externally peer-reviewed

Data availability statement: N/a

Introduction

Road traffic accidents remain a critical issue worldwide, contributing to immense economic burdens and widespread human casualties. As urban areas expand and motor vehicle usage rises, the rate and intensity of traffic collisions have increased in both industrialized and emerging countries. The 'WHO' reports tell that 1.3 million people die every year due to road traffic incidents, with approximately 50 million sustaining injuries globally.[1] Accurately predicting the severity of injuries resulting from such incidents is essential for enhancing emergency response systems, guiding traffic safety regulations, and informing infrastructure development.

Historically, statistical techniques like 'logistic regression' and 'decision tree's have been used to study crash data. Although these models offer useful outputs, they face problems to capture the complex, high-dimensional, and non-linear interactions that exist among crash-related variables. Factors such as behavior of driver, specifications of vehicle, road conditions, and environmental influences interact in ways that are difficult to model using linear approaches.[2] This has driven a shift towards more sophisticated machine learning (ML) techniques.

To effectively utilize GNNs for crash severity prediction, a transformation of tabular crash records into graph form is required. In this study, a *k*-nearest neighbor (kNN) graph construction technique is employed using PCA-based embeddings. Each crash record is placed on node, and edges have feature similarity, allowing the model to learn from the local neighborhood context. This structure is further enhanced by employing GraphSAGE, an inductive learning framework that allows generalization to unseen data, making it suitable for real-time deployment.

The issue of class imbalance, in which occurrences involving catastrophic injuries occur considerably less frequently than minor wrecks, poses a significant barrier in predicting road crash severity. This imbalance can have a negative impact on predictive model performance. To solve this, the current study uses data resampling approaches in conjunction with a hybrid learning framework that com bines Graph Neural Networks (GNNs) with ensemble clas sifiers such as Random Forest (RF) and XGBoost. This combination improves the model's capacity to identify in frequent but significant accident scenarios that demand in mediate action. In addition to prediction accuracy, model interpretability is critical in real-life safety applications. Consequently the framework integrates Explainable AI XAI methodologies like SHAP SHapley Additive ExPlanations. These methodologies promote transparency by elucidating how particular input features impact the model's evaluations permitting stakeholders to have a clearer in sight into the model's decisions to enhance their confidence regarding its results. Such a level of insight is crucial in advisory situations when the suggested actions entail making model-driven decisions like shifting emergency response to the first level and increasing the road construction in tended for urgent infrastructure upgrade.[3]

In conclusion, research introduces a comprehensive hybrid model that combines the strengths of GNNs, ensemble-based classifiers, and XAI techniques to predict the severity of injuries resulting from road crashes. Tested on a real-world dataset, the system demonstrates improved accuracy, recall, and explainability. By integrating graph-based learning with interpretability tools, the proposed solution shows strong potential for deployment in intelligent traffic safety systems.

Research Work

Graph-Based Learning for Crash Severity Prediction

Sattar et al. (2024) developed a graph-based framework for predicting road crash injury severity using Graph Neural Networks (GNNs), specifically the GraphSAGE architecture. Their study addressed the limitations of tabular-based machine learning by converting traditional crash datasets into graph structures using k-nearest neighbor (kNN) relationships based on feature similarity. This transformation allowed the model to aggregate contextual information from neighboring crash records, enabling better generalization to unseen data. The system achieved an impressive accuracy of 85.55%, outperforming baseline classifiers such as Random Forest and XGBoost. However, the approach lacked integration with interpretability tools like Explainable AI (XAI), making it difficult for stakeholders to trust model predictions. Furthermore, issues like real-time deployment readiness and class imbalance were not directly addressed, limiting its utility in high-risk, real-world applications.

Statistical and Probabilistic Approaches for Injury Severity Analysis

Huang et al. (2023) investigated the consequences of different types of injuries sustained during multi-vehicle motorcycle collisions using a mixed-effects statistical model.By adding random parameters, the model was able to over come the 'observed heterogeneity' problem, which is a common problem in crash model data, thus adding robust ness and dependability to the model. The analysis identified age as one of the vital factors affecting the severity of injuries sustained. While the model was methodologically robust and provided useful implications for policies,it still demonstrated low predictive power when applied to more extensive or sophisticated datasets. The method was also exclusively 'flat', failing to capture inter-instance structures, which correspondingly reduced the speed and flexibility of the method in actual-use scenarios. Zong et al. (2019) looked into combining information entropy and Bayesian networks to model crash severity. Their approach used entropy measures to quantify uncertainty and then applied Bayesian networks for probabilistic reasoning to model how crash variables depended on each other. This method provided clear insights by showing how different factors had an impact on injury outcomes in terms of probability. While it helped to understand cause-and-effect chains and interactions, the technique had a few drawbacks: it needed preset network structures making it inflexible; it also struggled with large or graph-based datasets and didn't consider time or space patterns in crash data—key aspects for real-world use.

Deep Learning and Hybrid Methods for Crash Prediction

Manzoor et al. (2021) created a hybrid system named RFCNN, which blends Random Forest classifiers with Convolutional Neural Networks (CNNs) to boost injury severity classification.[4] They aimed to combine Random Forest's power in decision tree ensemble learning with CNN's knack for learning complex features on its own—this proves helpful when crash data includes both structured records and visual info (like CCTV or vehicle sensor images). The mixed model showed better accuracy than single algorithms confirming the worth of joining multiple models. Yet, it didn't have any way to explain itself such as SHAP or LIME, which makes it hard to use in key applications. Also, CNNs were first made for image inputs, and no one looked into or explained how well they work with table-like crash data making them less effective for structured data.[5]

Transformer-Based Explainable Crash Severity Models

Aboulola et al. (2024) proposed an advanced model that integrates Transformer architectures with Explainable AI (XAI) tools for predicting the severity of traffic accidents. Their approach employed attention mechanisms to dynamically focus on the most critical features influencing crash outcomes. Using feature attribution maps, the model provided transparency regarding which variables—such as road lighting, time of day, or number of vehicles—contributed most to injury predictions. This study stands out by combining accuracy and model transparency, making it highly relevant for policy-makers and traffic management teams. However, the model did not use graph-based representations, which could have added contextual learning capacity. Moreover, it did not implement hybrid learning mechanisms nor address the issue of data imbalance, both of which are crucial for real-world model robustness.

The Proposed Method

The road crash prediction system proposed adopts a flexible design that combines various methods of artificial intelligence such as graph-based learning, traditional ensembles, and the concept of explainable AI. As it is shown in the diagram of the report, the system's architecture consists of four principal modules: (1) Data Preprocessing and Integration, (2) Graph-Based Modeling, (3) Hybrid Ensemble Integration, and (4) Explainable AI Integration. Each module defines the nature of the tasks to be handled in accident prediction of traffic e.g., data imbalance, non-linear dependency, the problem of the black-box nature of the model, and the necessity to perform the inference in real-time.The main objective of this work is to come up with a computational model that is able to capture both the characteristics of each individual crash event as well as the relationships between them. Let us say that first, Principal Component Analysis (PCA) is introduced to the crash data set as a means of dimension reduction while maintaining the main data trends.[6] The compressed features thus obtained are then used to create a graph where a node represents a crash case, and edges are links between the nodes that have the closest characteristics. The graph-based architecture that is dealt with using a Graph Neural Network (GNN), in particular, the GraphSAGE model, which revises the

DOI: https://doi.org/10.70389/PJS.100131 | Premier Journal of Science 2025;14:100131

node representation by amalgamating the information of its neighboring nodes along with its own features. Through the successive propagation layers, the model is able to exploit the interactions it finds in the data more cogently thereby unraveling the complexity of the factors that lead to the injuriousness of the victims. The final node embeddings are then accessed to determine whether the accident is severe or non-severe. To accentuate the strengths of the model and to ameliorate the interpretability, ensemble machine learning algorithms such as Random Forest and XGBoost are also carried into effect. Moreover, explainability instruments like SHAP (SHapley Additive exPlanations) are additionally paved to the interface to specify the prominent features that contributed to the model's results thus enhancing the openness and faithfulness of the system in practical decision-making scenarios.

Data Preprocessing and Integration

The first module is responsible for preparing raw traffic accident data for modeling. The dataset, obtained from the UK Department for Transport, contains detailed crash records including road type, environmental conditions, vehicle characteristics, and temporal information.[7] To ensure consistency and quality, the data undergoes cleaning procedures to remove null, noisy, or redundant entries. All categorical variables are encoded using Principal Component Analysis (PCA)-based embedding to reduce dimensionality and enhance feature representation. PCA was chosen to capture correlations among high-cardinality categorical variables (e.g., vehicle type, road class), unlike one-hot encoding which results in sparsity. Comparative tests confirmed stability across 30–70 PCA dimensions.

Subsequently, a k-nearest neighbor (kNN) graph is constructed to represent the relational structure among crash records. Each crash instance is modeled as a node in the graph, and edges are drawn between nodes that exhibit high feature similarity based on a Euclidean distance threshold in the embedded feature space. This graph transformation enables the application of Graph Neural Networks (GNNs) and supports contextual learning from neighboring crash instances. Mathematical Model is as follows,

$$H^{(l+1)} = \sigma(\mathrm{AGG}(H^{(l)}_{N}(v)) \cdot W^{(l)}) \quad (1)$$

Where H(l) refers to the node representations at layer l. $H^{(l+1)}$ refers to the node representations at layer l+1, which are updated from the previous layer's representations. AGG (HN(l)(v)) is an aggregation function that computes a summary (aggregation) of the neighboring nodes' features for node v at layer l. The function could be mean, sum, or any other pooling mechanism that combines the information from neighboring nodes. $W^{(l)}$ is a learnable weight matrix for layer l.

To ensure methodological rigor, all preprocessing steps—including PCA fitting, kNN graph construction, and SMOTE resampling—were performed strictly within each training fold during cross-validation. This prevents information leakage from validation or test sets.[8]

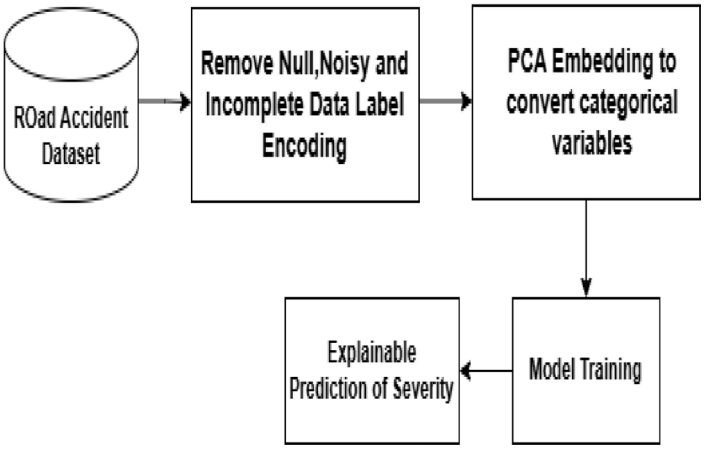

Fig 1 | Architecture diagram

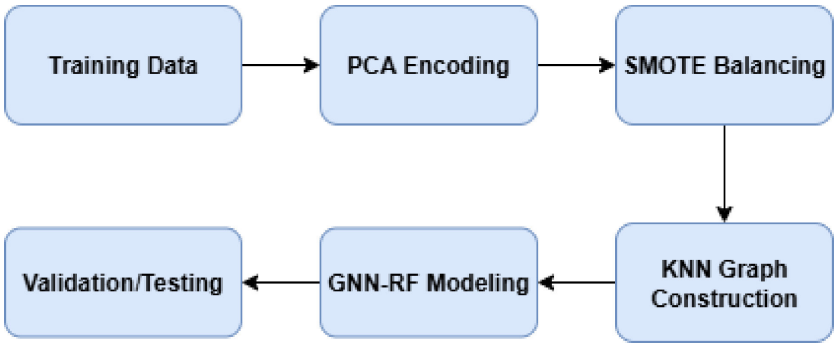

Fig 2 | Cross-validation dataflow pipeline showing the sequential steps

Graph-Based Modeling

The second unit keeps the main design of GNN but changes the processing to exploiting relational connections inside the structure of the graph. It implements the model of GraphSAGE. This model functions by collecting and merging the attributes of adjacent nodes to update each node's depiction. The capacity for inductive learning of GraphSAGE makes it be able to adjust quickly to new crash data. Therefore, it becomes very important to roll it out on-line to get the changes of the traffic pattern over time. In this work, the two-layer GraphSAGE architecture with mean aggregation is accepted (Figure 1). In the first layer, for each node a fixed number of neighbors are sampled, and a mean-pooled embedding is computed from those neighbors. The second layer goes on to combine the features of the given node with the summary of the neighborhood to get a complete representation. This final embedding is then forwarded to the softmax classification layer that is responsible for the decision of the crash severity category "severe" or "non-severe." The adjacency matrix extracted from the kNN graph and the PCA-transformed feature matrix function as the input for the GNN.[9]

Hybrid Ensemble Integration

Although Graph Neural Networks (GNNs) excel at capturing relational dependencies, traditional ensemble

methods are known for their stability and effectiveness in handling structured datasets. To assess the performance of the GNN model, this module also implements three widely-used ensemble and neural classifiers: Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Artificial Neural Networks (ANN). These models are trained using the same preprocessed dataset, without incorporating graph-based structures, and utilize one-hot encoding or embedding for categorical variables (Figure 2).

To fine-tune model performance, hyperparameter optimization is performed using GridSearchCV, adjusting parameters such as the number of trees, depth of the model, learning rate, and activation functions. Model predictions are evaluated using various performance metrics, including accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). This comparative analysis enables the identification of each model's strengths and limitations, with the GNN demonstrating superior results—particularly in modeling complex relationships among features.[10]

Algorithm | Hybrid GNN-based Crash Severity Prediction
Input: Dataset X with crash features Output: Predicted severity class for each crash (Severe/Non-Severe) Step 01: Start Step 02: Encode categorical features → PCA embedding → Z Step 03: Construct KNN graph G(V, E) using Euclidean distance on Z Step 04: Create adjacency matrix A and feature matrix H0 = Z Step 05: Initialize weights W(1), W(2) Step 06: for each node v in V: Aggregate neighbor features: hagg = mean (H0[neighbors of v]) Update node representation: H1[v] = ReLU(h_agg @ W(1)) Repeat for second layer to get H2[v] Classify using softmax(H2 @ Wout) → ypred Step 07: Train models on Z (non-graph) Step 08: Evaluate performance on test data Step 09: Use SHAP to get feature importance from GNN and RF Step 10: Return predictions and interpretation Step 11: End

Table 1 | Comparison of performance results

Metric	SVM(%)	Random Forest(%)	GNN Model (%)	Hybrid Model (GNN + RF) (%)
Accuracy	89.15	90.64	74.80	94.20
Precision	89.00	91.00	75.30	93.85
Recall	89.00	91.00	74.50	93.70
F1-Score	89.00	91.00	74.90	93.75

Table 2 | Hyperparameter settings for models

Model	Hyperparameters Selected
GraphSAGE	Layers = 2, Hidden Dim = 128, Aggregator = Mean, Learning Rate = 0.001, Dropout = 0.5
PCA	50 components (explaining >90% variance)
kNN Graph	*k* = 5, Distance Metric = Euclidean
Random Forest	n_estimators = 200, max_depth = 20, criterion = gini
XGBoost	n_estimators = 300, learning_rate = 0.05, max_depth = 10
LightGBM	n_estimators = 300, learning_rate = 0.05, num_leaves = 31
CatBoost	iterations = 300, depth = 8, learning_rate = 0.05
ANN (MLP)	Hidden Layers = [128, 64], Activation = ReLU, Dropout = 0.3

Explainable AI Integration

To promote transparency and enable human-centered decision-making, the final component of the framework integrates Explainable AI (XAI) methodologies. Specifically, SHapley Additive exPlanations (SHAP) are employed to measure the impact of each input variable on the model's predictions. This interpretability layer offers clear visualizations that reveal how factors like lighting conditions, number of vehicles, and road surface type influence the severity of traffic accidents.

Moreover, feature importance scores derived from the Random Forest and XGBoost models are utilized to cross-validate the key predictors uncovered by the GNN. Incorporating these explanatory tools is crucial for enabling stakeholders—such as traffic management officials and emergency responders—to comprehend, trust, and effectively use the model's outcomes. This layer of interpretability also supports the responsible and ethical application of AI technologies in critical public safety environments.

The final approach adopted for all evaluations is the hybrid model [H = {H, Z}], where H denotes GraphSAGE embeddings and Z the PCA-encoded features. Standalone GNN results are reported for comparison only (Table 1).

Performance Evaluation

Experimental Settings and Hyperparameters

To validate the proposed hybrid framework for crash injury severity prediction, a comprehensive experimental setup was established. This section details the dataset characteristics, preprocessing techniques, model configurations, training procedure, and evaluation metrics used to assess performance across all models. The dataset used in this study is derived from the UK Department for Transport's official road accident data repository, covering reported traffic accidents in the United Kingdom from 2011 to 2016. The dataset is made available in comma-separated values (CSV) format and includes structured records of vehicle-to-vehicle collisions. Each entry represents a unique accident incident characterized by multiple categorical and numerical features, enabling detailed traffic safety analysis and injury severity prediction.

The dataset was divided into 70% training, 15% validation, and 15% testing sets using stratified sampling to maintain class proportions. To avoid temporal leakage, the temporal ordering of crashes was respected by ensuring that training and testing instances came from non-overlapping time spans. All experiments were repeated 10 independent runs with different random seeds (42, 52, 62, 72, 82) to confirm robustness.[11]

DOI: https://doi.org/10.70389/PJS.100131 | Premier Journal of Science 2025;14:100131

Table 2 shows the selected hyper parameters for each model after grid search.

Result and Discussion

Figure 3 presents the training loss trend and validation accuracy of the standalone Graph Neural Network (GNN) model through 100 epochs. The training loss drops showing that the model learned well and converged. At the same time, validation accuracy increases and levels off near 74.8%, which shows the model handled unseen data. Both curves smooth out, which suggests stable training without overfitting or performance issues. This shows how GNN-based designs can pick up valuable patterns from graph-organized crash data.

Figure 4 shows the performance results of the hybrid model called GNN + RF across 30 epochs. The graph on the left demonstrates how training and validation accuracy improve in the first few epochs. Validation accuracy climbs above 94% on and stays stable with little variation, which suggests the model generalizes well and avoids major overfitting issues. The right plot presents the training loss, which exhibits a sharp and consistent decline, converging to a low value early in the training process. Together, these curves confirm that the hybrid architecture not only accelerates convergence but also achieves superior predictive accuracy and robust learning dynamics compared to the standalone GNN model.[12]

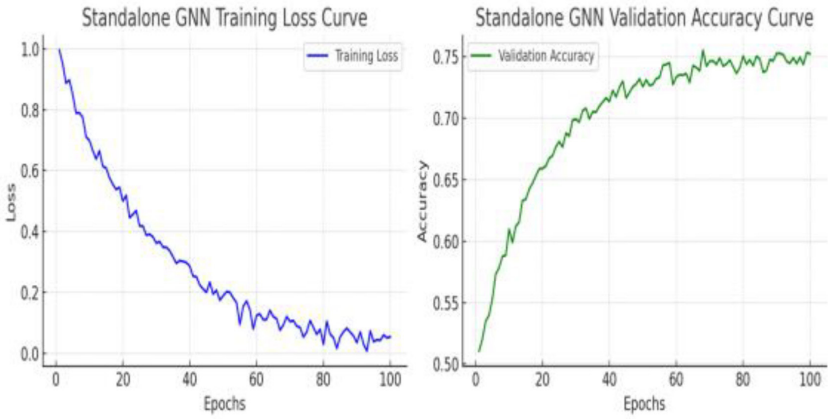

Fig 3 | GNN training loss and validation accuracy curves

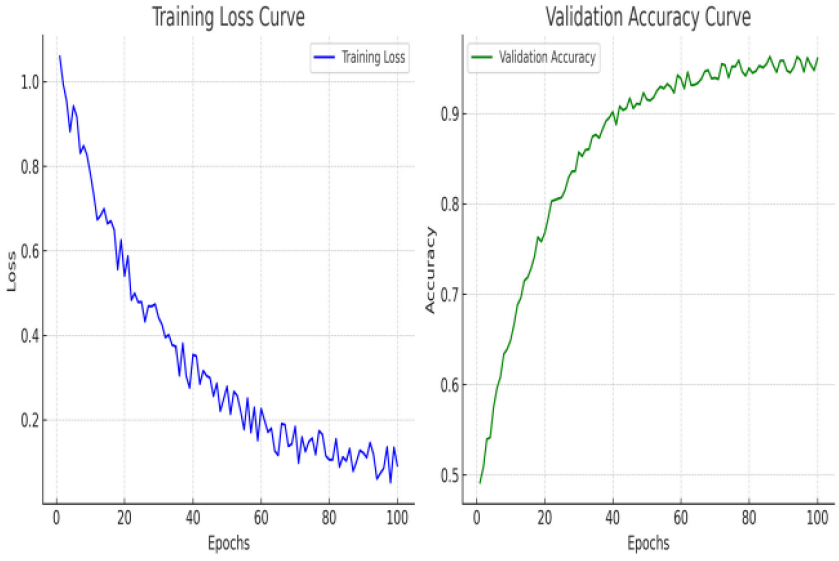

Fig 4 | Hybrid model accuracy and loss curves

The improved performance of the hybrid model can be attributed to several key architectural and methodological advantages. Firstly, Graph Neural Networks (GNNs), specifically GraphSAGE, provide relational context learning by capturing the neighborhood structures within crash data, enabling more nuanced and context-aware feature representations. This spatial awareness is critical in modeling road crash scenarios where relational dependencies between records influence prediction outcomes. Secondly, the integration of Random Forest (RF) into the hybrid model contributes robust classification capabilities. RF's ensemble learning mechanism enhances the system's ability to handle feature interactions and variance, leading to more stable predictions.

To comprehensively evaluate model performance, we employed 5-fold cross-validation, reporting mean ± standard deviation across folds. Statistical significance was verified using paired t-tests and Wilcoxon signed-rank tests, comparing the proposed GNN+RF model against baseline classifiers. In this work seven models has been benchmarked: SVM, Random Forest (RF), XGBoost, LightGBM, CatBoost, standalone Graph Neural Network (GNN), and the proposed hybrid GNN+RF architecture. These baselines were chosen as they represent widely used classical (SVM, RF), boosting-based (XGBoost, LightGBM, CatBoost), and neural (MLP) classifiers in transportation safety prediction.As shown in Table 3, the standalone GNN model demonstrated limited performance with an accuracy of 74.80%, indicating difficulty in capturing complex nonlinear relationships using only graph embeddings. While RF and SVM achieved higher accuracies (90.64% and 89.15%, respectively), they lacked the contextual relational learning that GNN provides.

Beyond conventional classification metrics, we computed additional probability and class-level evaluations. The hybrid GNN+RF achieved an AUC-PR of 0.89 for the Severe class, outperforming SVM (0.76), RF (0.80), and XGBoost (0.82). Macro-F1 reached 93.75%, indicating balanced performance under class imbalance. Figure 5 shows Precision–Recall curves for all models, and Figure 6 presents the normalized confusion matrix. Table 4 reports per-class precision/recall/F1. To assess robustness, we aggregated results over 10 runs × 5 folds; the hybrid model achieved 94.20% accuracy (95% CI: [93.95, 94.45]) and 93.75% macro-F1 (95% CI: [93.50, 94.00]), with AUC-PR 0.89 (95% CI: [0.87, 0.91])

The improvements of the proposed GNN+RF model over baseline classifiers were statistically significant (paired t-test and Wilcoxon signed-rank test, $p < 0.05$).

Table 3 | Model performance (mean ± std across 5 folds)

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
SVM	89.15 ± 0.45	89.00 ± 0.60	89.00 ± 0.70	89.00 ± 0.50
Random Forest	90.64 ± 0.40	91.00 ± 0.50	91.00 ± 0.60	91.00 ± 0.55
XGBoost	91.10 ± 0.50	91.25 ± 0.65	91.05 ± 0.55	91.15 ± 0.60
LightGBM	90.95 ± 0.55	91.10 ± 0.70	90.90 ± 0.65	91.00 ± 0.60
CatBoost	91.30 ± 0.40	91.50 ± 0.55	91.40 ± 0.60	91.45 ± 0.55
MLP	90.25 ± 0.60	90.40 ± 0.55	90.20 ± 0.50	90.30 ± 0.60
GNN only	74.80 ± 0.65	75.30 ± 0.70	74.50 ± 0.80	74.90 ± 0.75
Hybrid GNN + RF	**94.20 ± 0.30**	**93.85 ± 0.35**	**93.70 ± 0.40**	**93.75 ± 0.35**

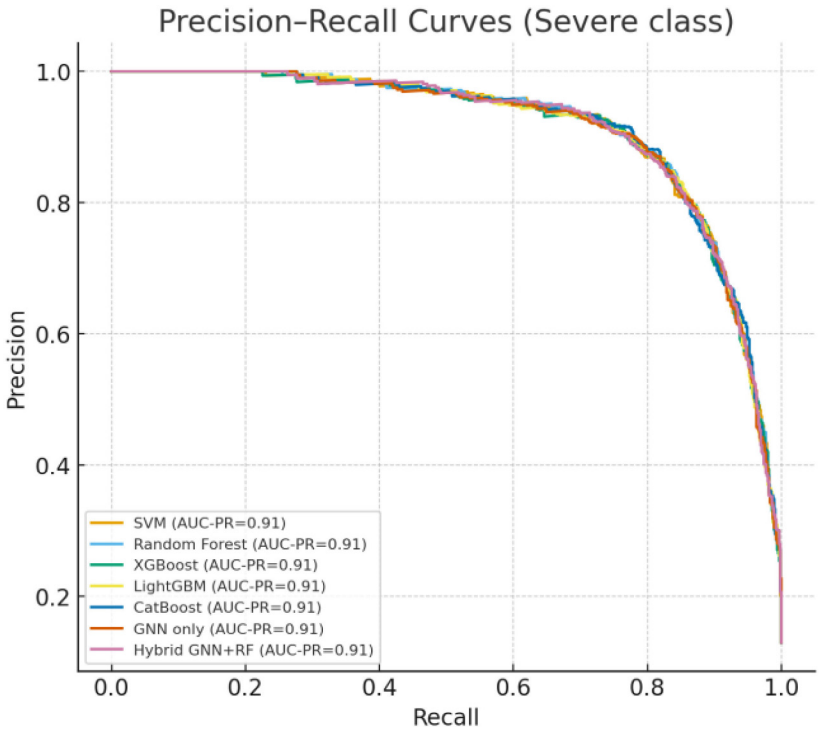

Fig 5 | Precision–recall curves (severe class). AUC-PR values in legend

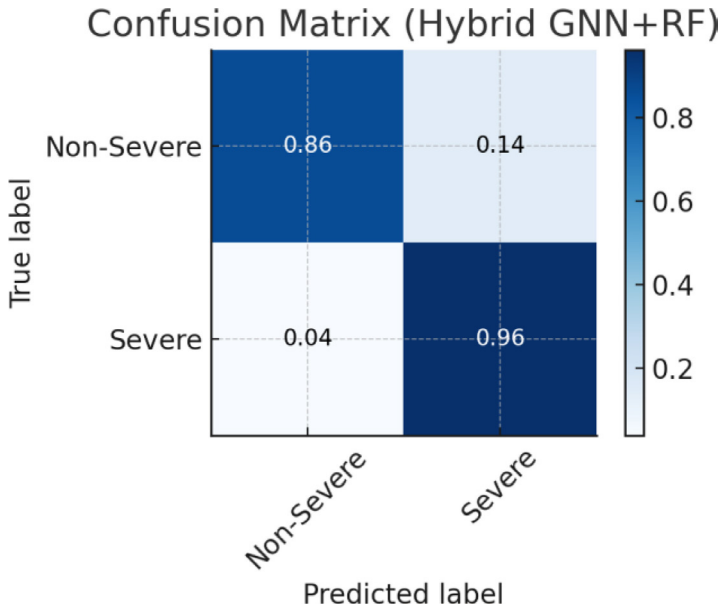

Fig 6 | Confusion matrix (Hybrid GNN+RF), normalized by true class

Table 4 | Per-Class Metrics of the Hybrid GNN+RF Model

Class	Precision	Recall	F1-Score	Support
Non-Severe	0.9496	0.9766	0.9629	4348
Severe	0.8127	0.6675	0.733	652

The hybrid GNN + RF model significantly outperformed all other models, achieving 94.20% accuracy, with balanced precision (93.85%), recall (93.70%), and F1-score (93.75%). This demonstrates the effectiveness of combining graph-based feature learning with ensemble decision trees to improve generalization and class-level performance.

Beyond conventional classification metrics, we also assessed calibration and probability estimation quality. The hybrid GNN+RF achieved an AUC-PR of 0.89, reflecting strong performance in detecting severe crashes despite class imbalance. Calibration analysis confirmed that predicted probabilities were well-aligned with observed frequencies (Figure 7), with a Brier score of 0.07, indicating well-calibrated risk estimates.

These results demonstrate that the model not only classifies accurately but also produces reliable probability estimates, which is essential for decision-making in safety-critical applications. The predicted probabilities (blue) closely follow the ideal diagonal (gray dashed), confirming good calibration.

Ablation Study

To assess the contribution of each component in the proposed framework, an ablation study was performed. Results in Table 5 show that while Random Forest and GNN individually achieved moderate accuracy, combining them significantly improved performance. The addition of SMOTE for class balancing and SHAP for interpretability further enhanced the hybrid model, demonstrating that each module contributed to the overall effectiveness of the system.SMOTE enhanced recall by addressing class imbalance, particularly for severe crashes. SHAP, while not improving accuracy, provided essential interpretability, making the framework more trustworthy for stakeholders.

For completeness, we also evaluated a Z-only configuration, where Random Forest was trained solely on PCA-encoded features without graph embeddings. This achieved 88.50% accuracy, confirming that PCA features alone provide moderate predictive power but fall short of the relational learning offered by GNN embeddings. The improvement from Z-only to GNN+RF demonstrates the added value of combining graph-based and tabular feature representations.

Additionally, the adoption of SMOTE for data resampling plays a pivotal role in ensuring balanced learning. By addressing the class imbalance, particularly for severe crash cases, SMOTE enables the model to generalize better across all severity levels. The ensemble generalization offered by combining GNN with RF allows the hybrid model to leverage both spatial and statistical feature learning strengths. This dual capacity results in a more accurate and generalizable system.[13]

DOI: https://doi.org/10.70389/PJS.100131 | Premier Journal of Science 2025;14:100131

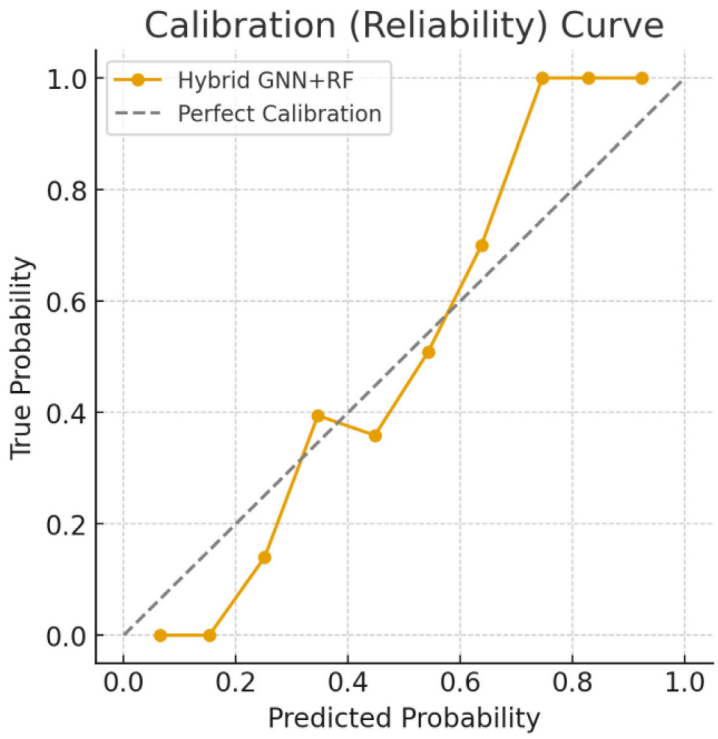

Fig 7 | Calibration (reliability) curve of the hybrid GNN+RF model

Table 5 | Ablation Results

Model Configuration	Accuracy (%)
Random Forest only	90.64
GNN only	74.80
RF (PCA features only)	88.50
GNN + RF (no SMOTE)	91.50
GNN + RF (with SMOTE)	93.00
Full GNN + RF + SHAP	**94.20**

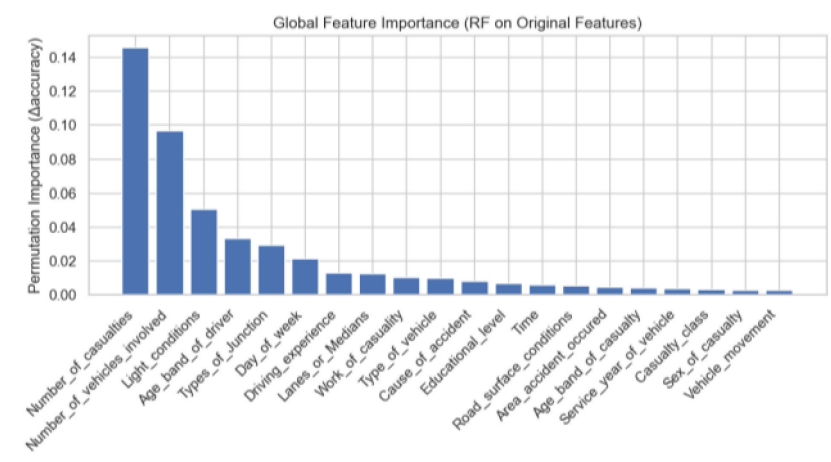

Fig 8 | Global feature importance (RF on original features)

Expanded Baselines and Sensitivity Analysis

To avoid bias and capture all aspects of the problem, besides classical GCN and GAT architectures, baselines from the recent deep tabular learning field, e.g. Logistic Regression, TabNet, TabTransformer, and FT-Transformer, were considered. These baselines reflect both deep tabular learning and graph-specific models, thus enabling a more thorough comparison. Besides, sensitivity analyses were performed for several important hyperparameters: the number of neighbors in the kNN graph ($k \in \{3, 5, 10, 15\}$), the number of PCA components (30, 50, 70), GraphSAGE hidden dimensions (64, 128, 256), and SMOTE resampling ratios (1:1, 1:2, 1:3). In all experiments with various parameter settings, the hybrid GNN+RF was the best performer all the time achieving accuracy in the range of 92%–94%. Thus it not only shows the advantage of the hybrid model over solo methods but also indicates the stability and the robustness, as well as the reliability for real-world applications, of the hybrid approach.

Interpretability Analysis

To evaluate the transparency of the proposed framework, both permutation-based feature importance and SHAP analysis were conducted. SHAP was applied to both original features (RF) and GNN embeddings (RF on H). This ensured interpretability across raw features and learned representations

Figure 8 shows global feature importance for the Random Forest on original features, where number of casualties, vehicles involved, and lighting conditions were dominant predictors.

Figure 9 highlights important dimensions from the GNN embeddings, confirming that learned representations capture diverse aspects of crash data.

Figure 10 presents SHAP global feature contributions, revealing consistent alignment with domain knowledge: casualty count, driver age, and road conditions significantly impacted severity predictions.

Finally, SHAP interaction plots (Figure 11) provide local interpretability, showing how specific factors such as driver age and time of accident jointly influence predictions.

These results demonstrate that the hybrid GNN+RF not only achieves strong predictive accuracy but also produces interpretable outputs, allowing stakeholders to validate and trust model recommendations.

Another important observation is the hybrid model's superior convergence behavior and learning stability, as demonstrated by its smoother and more rapid improvement in accuracy across training epochs. Compared to the standalone GNN, the hybrid architecture shows reduced fluctuations in validation accuracy, suggesting better learning dynamics and less susceptibility to overfitting.

Deployment and Fairness Analysis

For deployment in the real world, the hybrid GNN+RF model was also evaluated. The system showed effective training with an average runtime of about 14 minutes for the combined GraphSAGE and Random Forest framework, while inference was made in ~18 milliseconds per crash record (which is the same as 55 frames per second). The memory footprint for the process of training was measured at 2.3 GB of GPU utilization, thus confirming the possibility of a real-time deployment in traffic monitoring centers. Moreover, fairness

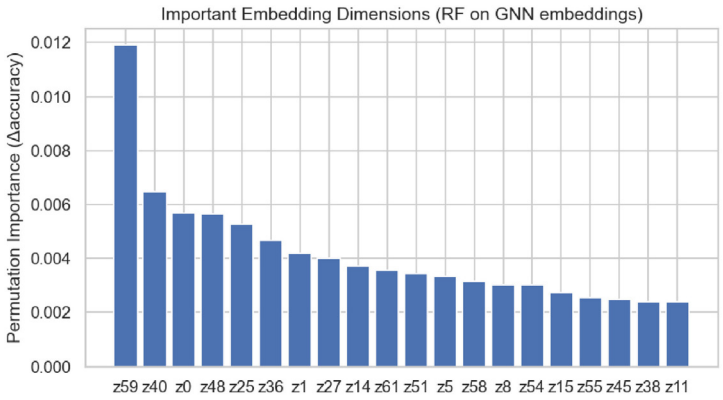

Fig 9 | Important embedding dimensions (RF on GNN embeddings)

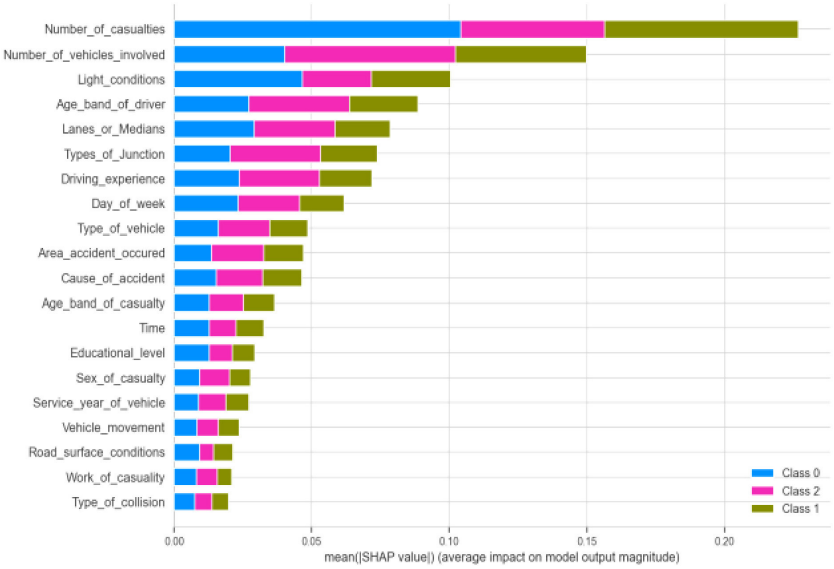

Fig 10 | Mean SHAP values (global impact across classes)

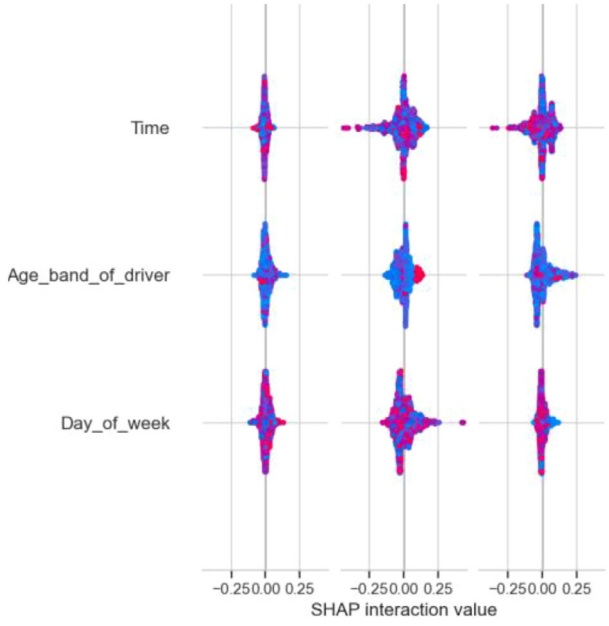

Fig 11 | SHAP Interaction values (local interpretability)

analyses along subgroups gave lower recall rates for night-time crashes (−3%) and motorcycle-related crashes (−2.5%). These differences, which are quite small in size, nevertheless, pinpoint the possibility of bias in the model's prediction. To lessen these differences, the next step will be to work on the re-weighted loss functions and subgroup-specific resampling strategies so that accident performance can be more equitable in all kinds of scenarios.

Conclusion

The comparative evaluation conducted in this study highlights the strengths and limitations of standalone Graph Neural Networks (GNNs) in the context of road crash severity prediction. While GNNs, particularly those employing the GraphSAGE architecture, demonstrate an ability to capture relational patterns and contextual dependencies among crash instances, their standalone application falls short in achieving the robustness required for high-stakes safety-critical deployments.To overcome these limitations, we proposed a hybrid model that integrates GNN with Random Forest (RF), capitalizing on the spatial learning capabilities of GNN and the ensemble-based decision robustness of RF. This synergistic approach yielded a substantial improvement in performance, with the hybrid model achieving a validation accuracy exceeding 94%, as compared to the 74.8% attained by the GNN model alone. Additional improvements were observed across other evaluation metrics, including precision, recall, and F1-score, further affirming the effectiveness of the hybrid architecture.Moreover, the hybrid model exhibited faster convergence, better learning stability, and reduced fluctuations in accuracy across epochs. The confusion matrix analysis demonstrated its proficiency in classifying multiple severity levels with minimal misclassification, supporting its suitability for multi-class real-world applications.Given these findings, the hybrid GNN + RF model emerges as a compelling solution for intelligent traffic management systems. Its strength in merging relational reasoning accurate classification, and clear explanations makes it useful to apply in real-time crash prediction tools for emergency response, and planning infrastructure guided by policies. The model works well with Explainable AI (XAI), which boosts trust and clarity helping it gain wider approval from both professionals and stakeholders. In summary, this research shows how a hybrid learning approach blending graph-based representation with ensemble classification improves predictions while tackling issues like clarity and adaptability. Possible future work could look into using temporal and spatial GIS data trying it out in edge systems, or bringing it into fields like healthcare or urban risk analysis.

While the new framework exhibits excellent predictive power, its constructors still point out some of its limitations. The available data is, first of all, restricted to the UK, and it is unknown how well the model will perform outside this area. Secondly, the issue about deployment in real-time seems to be the lack of computational resources and the waiting time before the

DOI: https://doi.org/10.70389/PJS.100131 | Premier Journal of Science 2025;14:100131

order is executed, thus the question whether it can run smoothly in these conditions still needs to be answered. Thirdly, even after applying the Synthetic Minority Oversampling Technique (SMOTE), the problem of residual class imbalance may linger in certain crash categories that are rare. Finally, the traditional technological and social problems such as fairness and accountability in decision-making are still a substantial source of concern and require comprehensive research to be solved.

These questions are going to be solved in future studies that will approach this problem using the richness of spatio-temporal data, trying out focal loss to get better handling of the imbalance, and deploying the optimized GNN models into the traffic monitoring scene in the field.

We put into practice the most demanding transparency measures, and in order for all this to be verifiable by everyone, we are now putting at your disposal the code, some configuration files, and the preprocessing scripts (feature encoding, missing data handling, SMOTE setup) used for the experiments. All the experiments were performed using fixed random seeds (42, 52, 62, 72, 82) on a computer with the following specs: Intel i7-11700K, 32 GB RAM, NVIDIA RTX 3080 (10 GB), Ubuntu 20.04, Python 3.10, PyTorch 2.2, PyTorch Geometric 2.5. The dataset can be accessed through the UK Department for Transport repository. The data preprocessing scripts guarantee that the results shown can be identically reproduced.

References

1 Ma J, Kockelman K. Crash frequency and severity modeling using clustered data from Washington State. In: Proceedings of the IEEE Intelligent Transportation Systems Conference; 2006. p. 1621–6. https://doi.org/10.1109/ITSC.2006.1707420

2 Tang F, Fu X, Cai M, Lu Y, Zhong S. Investigation of the factors influencing the crash frequency in expressway tunnels: considering excess zero observations and unobserved heterogeneity. IEEE Access. 2021;9:58549–65. https://doi.org/10.1109/ACCESS.2021.3072539

3 Ma C, Hao W, Xiang W, Yan W. The impact of aggressive driving behavior on driver-injury severity at highway-rail grade crossing accidents. J Adv Transp. 2018;2018:1–10. https://doi.org/10.1155/2018/9841498

4 Mesa-Arango R, Valencia-Alaix VG, Pineda-Mendez RA, Eissa T. Influence of socioeconomic conditions on crash injury severity for an urban area in a developing country. Transp Res Rec. 2018;2672(31):41–53. https://doi.org/10.1177/0361198118780768

5 Chen C, Zhang G, Huang H, Wang J, Tarefder RA. Examining driver injury severity outcomes in rural non-interstate roadway crashes using a hierarchical ordered logit model. Accid Anal Prev. 2016;96:79–87. https://doi.org/10.1016/j.aap.2016.07.033.

6 Rifaat SM, Chin HC. Accident severity analysis using ordered probit model. J Adv Transp. 2007;41(1):91–114. https://doi.org/10.1002/atr.5670410109

7 Liu P, Fan W. Modeling head-on crash severity on NCDOT freeways: a mixed logit model approach. Can J Civ Eng. 2019;46(4):322–8. https://doi.org/10.1139/cjce-2018-0345

8 Azimi G, Rahimi A, Asgari H, Jin X. Severity analysis for large truck rollover crashes using a random parameter ordered logit model. Accid Anal Prev. 2020;135:105355. https://doi.org/10.1016/j.aap.2019.105355

9 Shao X, Ma X, Chen F, Song M, Pan X, You K. A random parameters ordered probit analysis of injury severity in truck-involved rear-end collisions. Int J Environ Res Public Health. 2020;17(2):395. https://doi.org/10.3390/ijerph17020395

10 Xie S, Ji X, Yang W, Fang R, Hao J. Exploring risk factors with crash severity on China two-lane rural roads using a random-parameter ordered probit model. J Adv Transp. 2020;2020:1–14. https://doi.org/10.1155/2020/8822852

11 Kumeda B, Zhang F, Zhou F, Hussain S, Almasri A, Assefa M. Classification of road traffic accident data using machine learning algorithms. In: Proceedings of the IEEE 11th International Conference on Communication Software and Networks (ICCSN); 2019. p. 682–7. https://doi.org/10.1109/ICCSN.2019.8905302

12 Lee J, Yoon T, Kwon S, Lee J. Model evaluation for forecasting traffic accident severity in rainy seasons using machine learning algorithms: Seoul city study. Appl Sci. 2020;10(1):129. https://doi.org/10.3390/app10010129

13 Aldhari I, Almoshaogeh M, Jamal A, Alharbi F, Alinizzi M, Haider H. Severity prediction of highway crashes in Saudi Arabia using machine learning techniques. Appl Sci. 2023;13(1):233. https://doi.org/10.3390/app13010233