Predicting and Explaining Fatal Road Casualty Types in Great Britain: A Comparative Analysis of Machine Learning, Deep Learning, and Transformers

Umar Hasan[1][0009-0009-6659-2006] and Mohammad Abdul Qayum[1][0000-0002-4434-4345]

Department of Electrical and Computer Engineering
North South University, Dhaka 1229, Bangladesh
{umar.hasan, mohammad.qayum}@northsouth.edu

Abstract. Road accidents pose a significant global health and economic challenge. This research conducts a comparative analysis of machine learning (ML), deep learning (DL), and Transformer-based natural language processing (NLP) techniques to predict the specific type of fatal casualty in British road accidents, aiming to enhance understanding and prevention efforts. The study utilizes two correlated datasets detailing fatal road accidents and associated casualties in Great Britain from 2006 to 2008, sourced from data.gov.uk. Data preprocessing included exploratory data analysis, handling missing values, feature engineering, and correlation analysis. For traditional ML models (Logistic Regression, k-Nearest Neighbors, Decision Tree, Random Forest, XGBoost, CatBoost), class imbalance was addressed using SMOTE, and models were tuned using Randomized Search. For DL (TabNet, MLP) and Transformer models (BERT, RoBERTa, DistilBERT), a standard train/validation/test split was used without SMOTE, and tabular data was converted into descriptive text sequences for Transformer input. Model performance was assessed using accuracy, F1-score, the area under the receiver operating characteristic curve, and confusion matrices on a held-out test set. Explainability for the best models was explored using LIME. Results indicate that tuned gradient boosting models (Random Forest, XGBoost) achieved the highest accuracy (approx. 87%). Our research highlights the capabilities and trade-offs of diverse modeling approaches for identifying risk factors associated with fatal casualty types, informing targeted road safety strategies.

Keywords: Road Safety · Fatal Casualties · Machine Learning · Deep Learning · Transformers · XGBoost · Random Forest · TabNet · MLP · BERT · LIME.

1 Introduction

Road accidents represent a critical global public health issue, responsible for a substantial number of fatalities and severe injuries annually. The World Health

Organization (WHO) highlighted the severity of this problem, noting the significant human cost and projecting road traffic accidents (RTAs) as a potentially leading cause of death in the coming years [32]. Beyond the human tragedy, RTAs impose a considerable economic burden globally [32]. Understanding the intricate factors contributing to these incidents, particularly those resulting in fatalities, is paramount for developing and implementing effective road safety strategies to mitigate risks and save lives.

Existing studies have increasingly employed statistical and machine learning (ML) techniques to analyze road accident data, yielding valuable insights into accident occurrence prediction, injury severity assessment, and identifying contributing factors [1,3,17]. Recent advancements have seen the application of more sophisticated methods, including deep learning (DL) architectures [24, 30] and explainable AI (XAI) techniques like LIME and SHAP [12,15,26] to unravel the complex, non-linear relationships inherent in accident data [15]. Furthermore, natural language processing (NLP) models, particularly Transformers [29], are emerging as powerful tools for analyzing textual data related to accidents. However, their application to structured tabular accident data has been explored less. While significant work has focused on predicting accident severity [3, 8, 17], predicting the specific type of fatal casualty remains a crucial area requiring further investigation using diverse modeling paradigms.

This study addresses this gap by performing a comparative analysis of tuned traditional ML models, DL models, and state-of-the-art Transformer models to predict the type of fatal casualty in British road accidents. Leveraging two comprehensive, real-world datasets from the UK government's open data portal covering fatal accidents from 2006 to 2008, we evaluate these distinct approaches. We introduce a method for Transformer models to convert structured tabular data into descriptive text sequences. Explainability for high-performing ML models is investigated using Local Interpretable Model-agnostic Explanations (LIME) [25]. The insights derived aim to assist policymakers and traffic authorities create more targeted interventions and preventive measures, potentially aiding emergency services in prioritizing rescue efforts. The key contributions of this study include:

- Conducting a rigorous quantitative comparison of tuned traditional ML, DL, and Transformer models for predicting fatal casualty types based on standard performance metrics.
- Implementing and evaluating a novel method of converting structured tabular accident data into serialized text sequences for analysis with Transformer-based NLP models.
- Applying explainability techniques (LIME) to the best-performing models to interpret their predictions and identify key contributing factors on an instance-level, supplementing the quantitative analysis with qualitative insights.

2 Related Work

2.1 ML Applications in RTA Analysis

RTAs inflict a significant human and economic toll worldwide [32]. Researchers have increasingly turned to ML as a powerful tool for forecasting and analyzing RTAs [1, 14]. Various studies have explored ML techniques, demonstrating promising results in predicting RTA occurrences and severity [3, 4, 17, 18]. The complexity of RTAs, often involving multiple converging factors, makes traditional linear analysis difficult, whereas ML models excel at capturing intricate, non-linear relationships [15]. Gradient boosting machines like XGBoost [9], along with Random Forest (RF) [6], are frequently employed due to their strong performance on tabular data [11, 33, 34]. While these studies establish the utility of ML, comparing traditional ML with newer DL and NLP approaches for specific tasks like fatal casualty type prediction offers further avenues for research.

2.2 DL, NLP, and Explainability in RTA Analysis

Beyond traditional ML, DL models like Multi-Layer Perceptrons (MLPs) [22] and specialized architectures for tabular data such as TabNet [2] have been applied to RTA analysis [24, 30]. These models can potentially capture deeper patterns but often require careful tuning and larger datasets. Concurrently, the field of XAI has gained prominence, with methods like LIME [25] and SHAP [26] being used to understand the predictions of complex models, identifying critical factors influencing accident outcomes [12, 15]. Furthermore, NLP, particularly with the advent of Transformer models like BERT [10], RoBERTa [19], and DistilBERT [27], offers the potential for analyzing textual accident reports or, as explored in this study, converting structured data into text for analysis [29]. While severity prediction remains a common focus [5, 8, 14], the application and comparison of tuned ML, DL, NLP (via text conversion), and XAI techniques specifically for predicting the type of fatal casualty represent a vital research direction addressed herein.

3 Methodology

This project employs a comparative approach, evaluating traditional supervised ML algorithms, DL models, and Transformer-based NLP models for the multi-class classification task of predicting fatal casualty types. The ML models include tuned versions of Logistic Regression [13], k-Nearest Neighbors [16], Decision Tree [28], Random Forest [6], XGBoost [9], and CatBoost [23]. The DL models are TabNet and an MLP. The Transformer models are BERT, RoBERTa, and DistilBERT. We utilized two datasets of fatal road accidents and casualties in Great Britain from 2006 to 2008. Data preprocessing involved cleaning, merging, feature engineering, correlation analysis, and specific handling for different model types. Model performance is evaluated primarily using accuracy, F1-score, ROC AUC, and confusion matrices on a held-out test set. LIME is used to explain predictions of selected ML models.

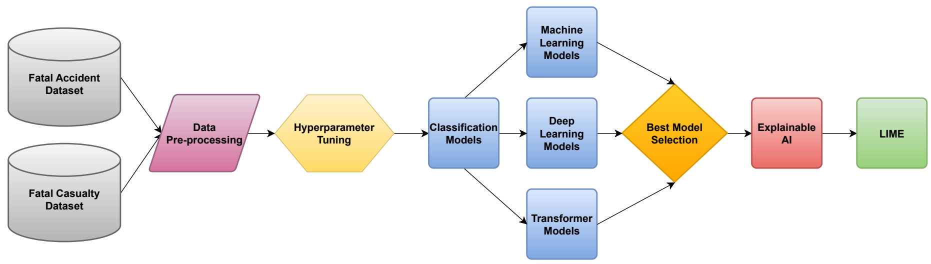


Fig. 1. Methodology Flowchart of the Study.

3.1 Dataset and Preprocessing

The datasets used are sourced from data.gov.uk: `fatalaccidentdata.csv` (7981 instances initially) and `fatalcasualtydata.csv` (8656 instances initially), covering 2006-2008.

Initial Cleaning and Merging. Rows with missing `Fatal_Accident_Index` were dropped from both datasets. The `Fatal_Casualty_Age` column was converted to a numeric type, and any rows with non-numeric or missing age values were subsequently dropped. The datasets were merged on `Fatal_Accident_Index` using an inner join, resulting in 8637 instances. Finally, any rows with a missing target variable (`Fatal_Casualty_Type`) were removed.

Exploratory Data Analysis (EDA). EDA was conducted to understand the dataset's underlying characteristics. A significant class imbalance was observed in the target variable, `Fatal_Casualty_Type`, as shown in Figure 3a, with 'Car Driver' being the predominant class. This finding directly motivated the use of the Synthetic Minority Over-sampling Technique (SMOTE) for the traditional ML models to ensure minority classes were adequately represented during training. Analysis of numerical features showed that casualty ages were most concentrated among young adults and older people. A feature correlation analysis (Figure 2) revealed only weak linear relationships, suggesting that non-linear models would be better suited to capture the complex patterns in the data.

Feature Engineering and Selection. Feature engineering was focused on preparing the data for modeling while retaining interpretability. The categorical `Fatal_Casualty_Sex` column was filtered to include only 'Male' and 'Female' entries, which were then converted into a new binary feature, `Fatal_Casualty_Sex_Binary`. For the ML and DL models, the numerical `Month_of_Accident` feature was one-hot encoded to better capture potential seasonal effects without imposing an ordinal relationship. These models' final features included accident time, location, vehicle counts, casualty counts, and specific casualty details like age and sex.

Data Splitting. *ML Models.* A 90% training and 10% testing split was used (Test set size: 864 instances), stratified by the target variable. SMOTE [7] was applied only to the training set to address the class imbalance identified during EDA (increasing the training sample size from 7773 to 33501), as illustrated

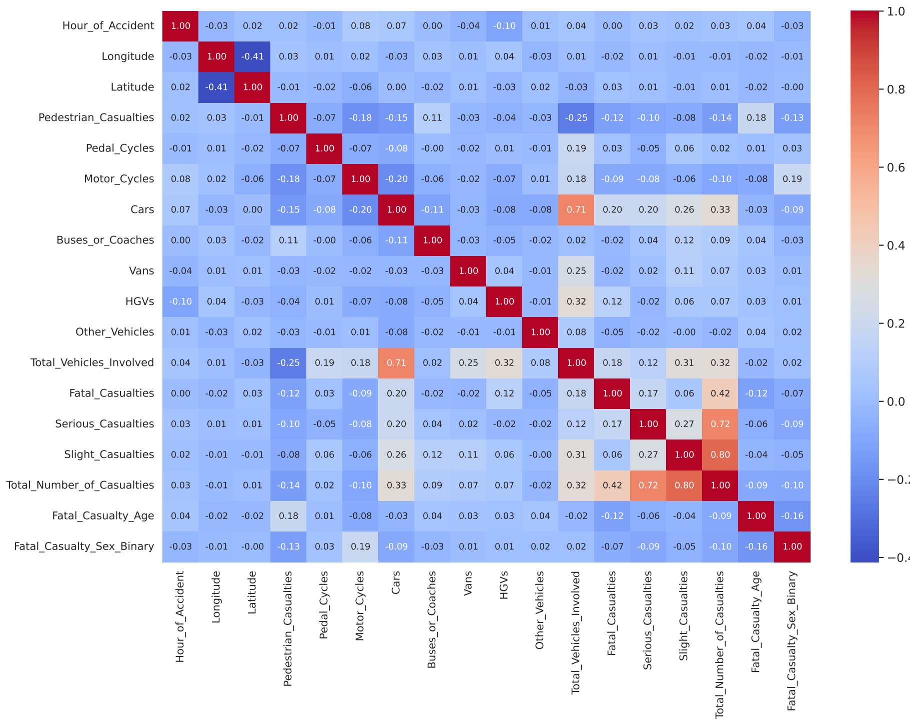
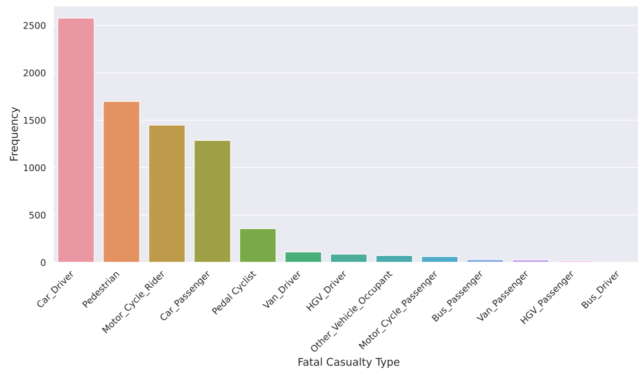


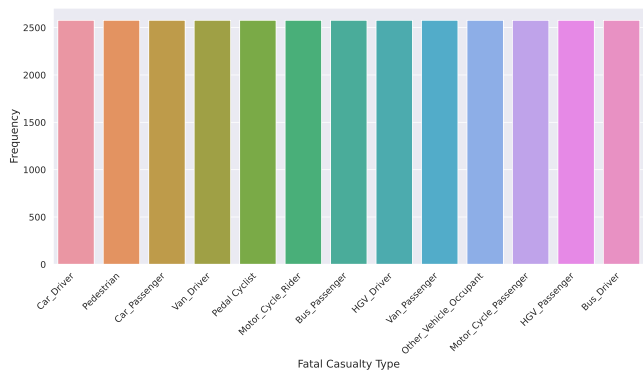
Fig. 2. Correlation Matrix of Features.

in Figure 3. Features were then scaled using a 'StandardScaler' fitted on the SMOTE-resampled training data. *DL and Transformer Models.* The data was split into training (70%, 6045 instances), validation (15%, 1296 instances), and test (15%, 1296 instances) sets, stratified by the target variable. This split was performed before any imputation or scaling to prevent data leakage. Missing values in features were imputed using the mean of the training set ('SimpleImputer'), and features were then scaled using a 'StandardScaler' fitted only on the imputed training data. SMOTE was not applied to these models.

Text Conversion for Transformers. For the Transformer models (BERT, RoBERTa, and DistilBERT), a template-based serialization method was used to convert each structured data row into a descriptive text sentence. This process concatenated feature names and their values into a human-readable format. For example, a row was transformed into a single string such as: *"Accident in Month 5, at hour 14, longitude -0.12, latitude 51.50; 2 vehicles involved (0 pedestrians, 0 pedal cycles, 1 motorcycles, 1 cars, 0 buses, 0 vans, 0 HGVs, 0 others); casualties: 1 fatal, 0 serious, 1 slight; casualty age 35, Male."* This serialized text became the direct input for the Transformer models, allowing them to process the tabular data as a sequence classification task.



(a) Training Set Target Class Distribution Without SMOTE.



(b) Training Set Target Class Distribution With SMOTE.

Fig. 3. Training Set Target Class Distribution.

Target Encoding. The categorical target variable `Fatal_Casualty_Type` (13 classes) was label encoded into a numerical format (0-12) for all models.

3.2 Machine Learning Architectures

Six traditional ML models were tuned using 'RandomizedSearchCV' with 5-fold cross-validation on the SMOTE-resampled, scaled training data. The best parameters are listed below.

Logistic Regression. '`C=0.047`', '`class_weight='balanced'`', '`l1_ratio=0.896`', '`max_iter=1335`', '`multi_class='multinomial'`', '`penalty=None`', '`solver='saga'`'.

k-Nearest Neighbors. '`metric='minkowski'`', '`n_neighbors=2`', '`p=1`', '`weights='distance'`'.

Decision Tree. '`criterion='entropy'`', '`max_depth=50`', '`max_features=None`', '`min_samples_leaf=6`', '`min_samples_split=19`'.

Random Forest. '`max_depth=30`', '`max_features='sqrt'`', '`min_samples_leaf=2`', '`min_samples_split=2`', '`n_estimators=293`', '`class_weight='balanced'`'.

XGBoost. `‘colsample_bytree=0.90’`, `‘gamma=0.35’`, `‘learning_rate=0.034’`, `‘max_depth=9’`, `‘n_estimators=189’`, `‘reg_alpha=0.61’`, `‘reg_lambda=0.85’`, `‘subsample=0.88’`.
CatBoost. `‘bootstrap_type=‘Bernoulli’’`, `‘border_count=128’`, `‘depth=9’`, `‘iterations=1086’`, `‘l2_leaf_reg=3.66’`, `‘learning_rate=0.077’`, `‘subsample=0.88’`.

3.3 Deep Learning Architectures

Two DL models were trained using PyTorch [21] on the scaled train/validation/test split without SMOTE.
TabNet. An attentive transformer network for tabular data [2]. Used parameters: `‘n_d=16’`, `‘n_a=16’`, `‘n_steps=4’`, `‘gamma=1.3’`, optimizer=‘AdamW’, `‘lr=2e-2’`, scheduler=‘ReduceLROnPlateau’, `‘mask_type=‘sparsemax’’`. Trained with early stopping.
Multi-Layer Perceptron. A feedforward neural network with two hidden layers (128 and 64 neurons), ReLU activations, BatchNorm, and Dropout (`rate=0.3`). Trained using AdamW optimizer (`‘lr=1e-3’`) and ‘CrossEntropyLoss’, with early stopping based on validation loss.

3.4 Transformer Architectures

Three pre-trained Transformer models from Hugging Face [31] were fine-tuned for sequence classification on the serialized text data: BERT [10], RoBERTa [19], and DistilBERT [27]. Models were fine-tuned for three epochs using the ‘Trainer’ API, with evaluation on the validation set guiding model selection.

3.5 Explainable AI

LIME was used to explain individual predictions of the tuned XGBoost and Random Forest models. LIME perturbs the input features of a single instance and fits a local, interpretable model to approximate the complex model’s behavior, highlighting the most influential features for that prediction.

4 Experiment and Results

4.1 Evaluation Metrics

The primary metrics used to evaluate and compare the models on the test set are:
Accuracy. The proportion of correctly classified instances.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

Table 1. Test Set Performance Comparison of All Models.

Model Family	Model Name	Accuracy	F1-score	Precision	Recall	ROC AUC
Machine Learning (with SMOTE)	**Random Forest**	**0.8704**	**0.8651**	**0.8651**	**0.8704**	**0.995**
	XGBoost	**0.8704**	**0.8653**	**0.8653**	**0.8704**	**0.995**
	CatBoost	0.8553	0.8510	0.8510	0.8553	N/A[a]
	Decision Tree	0.8426	0.8430	0.8451	0.8426	N/A[a]
	Logistic Regression	0.8009	0.8151	0.8551	0.8009	N/A[a]
	k-Nearest Neighbors	0.7778	0.7710	0.7695	0.7778	N/A[a]
Deep Learning (no SMOTE)	MLP	0.8673	0.8646	0.8636	0.8673	N/A[a]
	TabNet	0.8071	0.7889	0.7878	0.8071	N/A[a]
Transformers (no SMOTE)	BERT (bert-base-uncased)	0.8495	0.8452	0.8469	0.8495	N/A[a]
	DistilBERT (distilbert-base-uncased)	0.8441	0.8366	0.8350	0.8441	N/A[a]
	RoBERTa (roberta-base)	0.8279	0.8094	0.7945	0.8279	N/A[a]

[a]ROC AUC was computed only for the top-performing models (RF and XGBoost) for a focused comparison of their class discrimination capabilities.

F1-score. The harmonic mean of precision and recall, weighted by the number of true instances for each class (support).

$$\text{F1-score} = \sum_{l \in L} w_l \times \frac{2 \times (\text{Precision}_l \times \text{Recall}_l)}{\text{Precision}_l + \text{Recall}_l} \tag{2}$$

where L is the set of labels, and w_l is the proportion of instances belonging to label l.

ROC AUC (Micro-Average). Evaluates the model's ability to distinguish between classes across all thresholds.

Confusion Matrix. A table visualizing classification performance, showing counts of true and false predictions for each class.

4.2 Results and Analysis

This section presents the comparative performance of the tuned ML, DL, and Transformer models on their respective held-out test sets.

Table 1 summarizes the final test set performance. Key observations include:

ML Models. The tuned Random Forest and XGBoost models demonstrated the best performance among all models, achieving identical high accuracy (87.04%) and F1-scores (≈0.87) on their test set (N=864). Their strong performance is corroborated by excellent micro-averaged ROC AUC scores of 0.995 (Figure 7).

DL and Transformer Models. The MLP model performed competitively (86.7% accuracy) on its test set (N=1296) without SMOTE. Fine-tuned Transformer models also achieved good results, with BERT leading this group (85.0% accuracy).

Confusion Matrix Analysis. Detailed confusion matrices for RF (Figure 5) and XGBoost (Figure 6) show high accuracy for major classes like 'Car Driver' and 'Pedestrian'. However, they also revealed confusion between 'Car Driver' and 'Car Passenger', and they struggled with rare classes like 'HGV Passenger'.

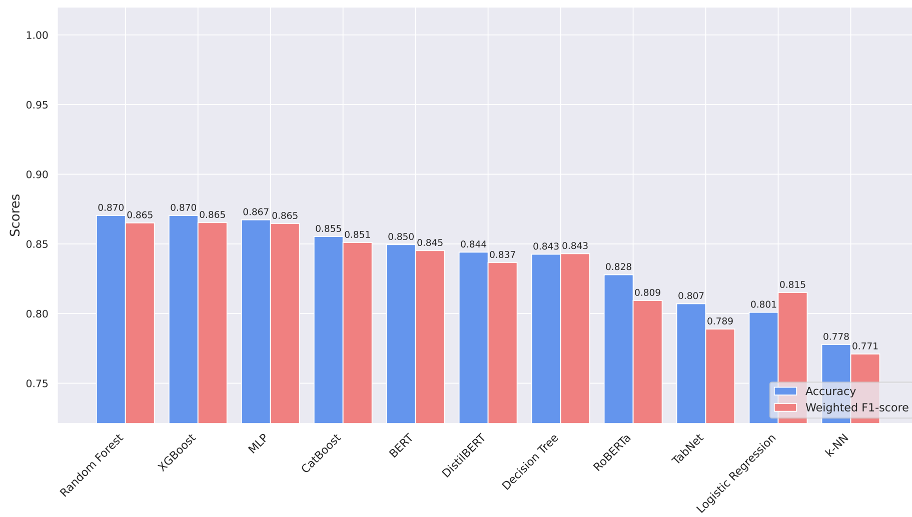


Fig. 4. Performance Comparison for All Models (Accuracy and Weighted F1-score).

Imbalance Handling (SMOTE). The application of SMOTE for ML models like Random Forest and XGBoost proved crucial, allowing them to learn better representations of minority classes and contributing significantly to their strong test set performance.
Explainability (LIME). LIME analysis for RF (Figure 8a) and XGBoost (Figure 8b) highlighted the local importance of features for specific predictions. For instance, for a prediction of 'Car Driver', both models identified low counts for other vehicle types (e.g., `Pedal_Cycles`, `Motor_Cycles`) as key supporting factors. This reinforces the need for XAI tools when deploying complex predictive models in safety-critical domains [15, 25].

Overall, the results suggest that well-tuned gradient boosting models (RF, XGBoost) trained with appropriate imbalance handling excel at this task. DL models like MLP and Transformers processing serialized text also achieve competitive performance.

4.3 Discussion

The experimental results demonstrate the effectiveness of various modeling approaches for predicting fatal road casualty types. Tuned ensemble ML models, particularly Random Forest and XGBoost, showed top performance, aligning with findings in similar RTA studies. Their ability to handle complex, non-linear interactions in the data was evident. A standard MLP also achieved strong results, suggesting neural networks can effectively model the data's inherent distribution. The successful application of Transformer models, using text converted from structured data, highlights a viable alternative approach, leveraging NLP

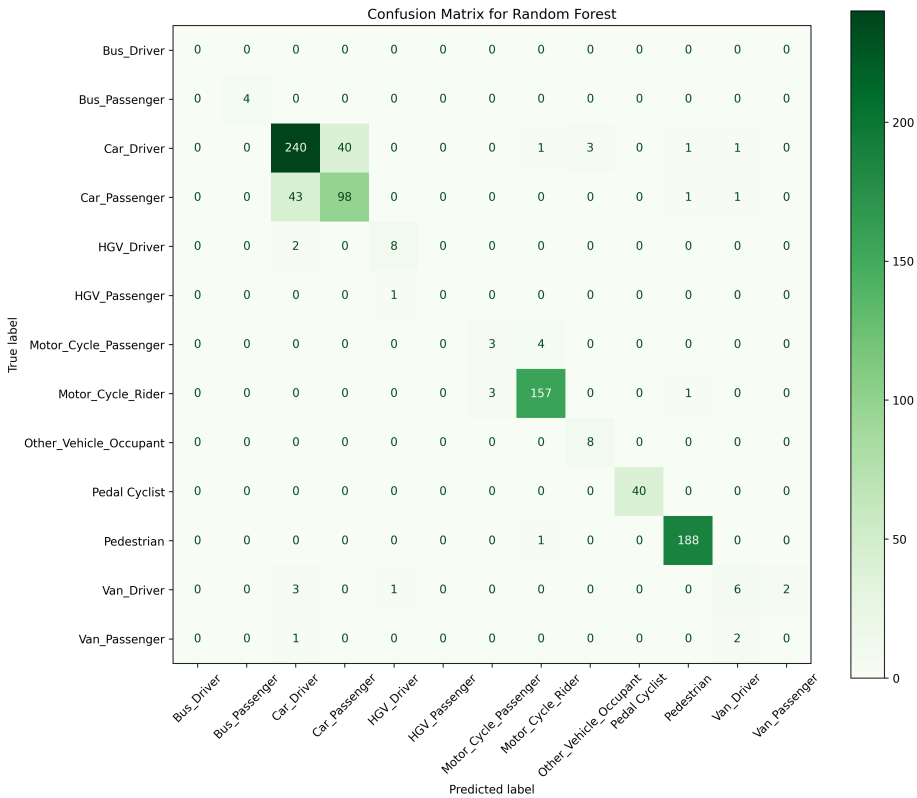


Fig. 5. Confusion Matrix for Random Forest.

capabilities for this classification task. While overall performance was high for the best models, challenges remain in accurately predicting rare casualty types.

Limitations. This study has several limitations: *Data Temporality and Scope.* The 2006-2008 dataset from Great Britain may not fully reflect current conditions or be directly generalizable elsewhere [20, 35]. *Feature Availability.* The dataset might lack granular details (e.g., driver behavior, precise weather) that could improve accuracy. *Model Interpretability.* While LIME provided local explanations for ML models, interpreting DL and Transformer models remains challenging. *Data Handling Discrepancies.* Differences in test set sizes and applying SMOTE only to ML models affect direct comparisons across model families regarding imbalance handling. These limitations highlight avenues for future research.

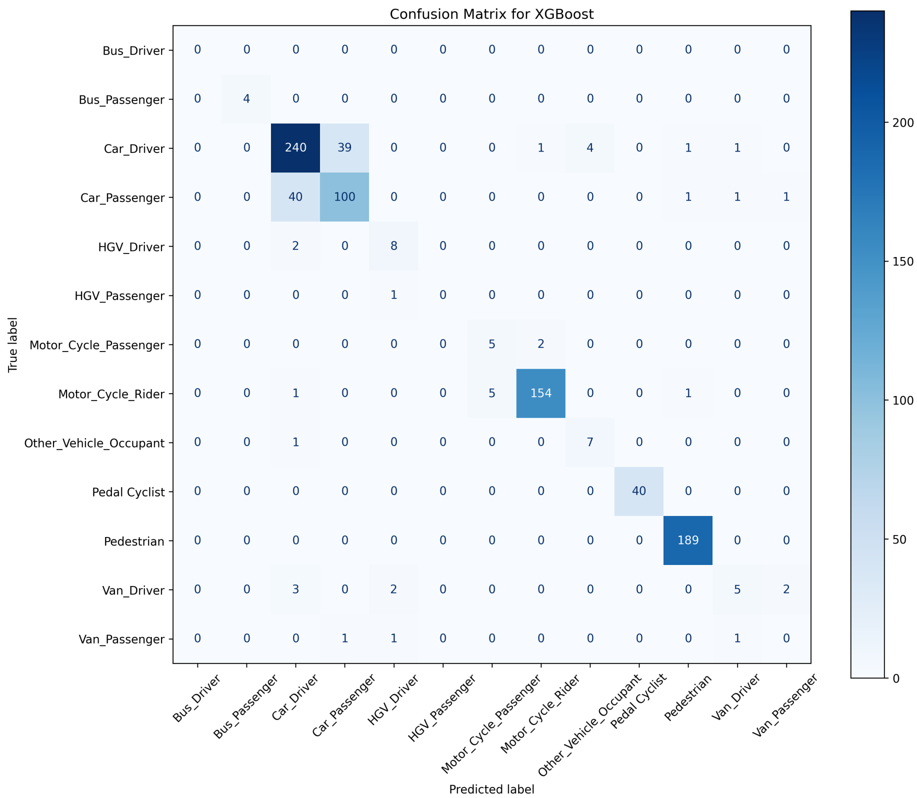


Fig. 6. Confusion Matrix for XGBoost.

5 Conclusion

This research successfully compared the performance of tuned traditional ML, DL, and Transformer-based NLP models to predict fatal road accident casualty types in Great Britain. By leveraging real-world accident data from 2006-2008 and employing techniques like SMOTE for ML model training, text conversion for Transformers, correlation analysis, ROC AUC evaluation, confusion matrix inspection, and LIME for explainability, we gained comprehensive insights into the capabilities of different approaches.

Our findings indicate that well-tuned ensemble ML models, specifically Random Forest and XGBoost trained with SMOTE, achieved the highest predictive accuracy (approx. 87%) and excellent ROC AUC scores (0.995) on their test set. A standard MLP also demonstrated strong, competitive performance (approx. 87% accuracy) on its respective test set without explicit oversampling. Transformer models (particularly BERT) processing textual representations of the data achieved respectable results (85% accuracy). Confusion matrices provided detailed insights into class-specific performance, highlighting common confusion

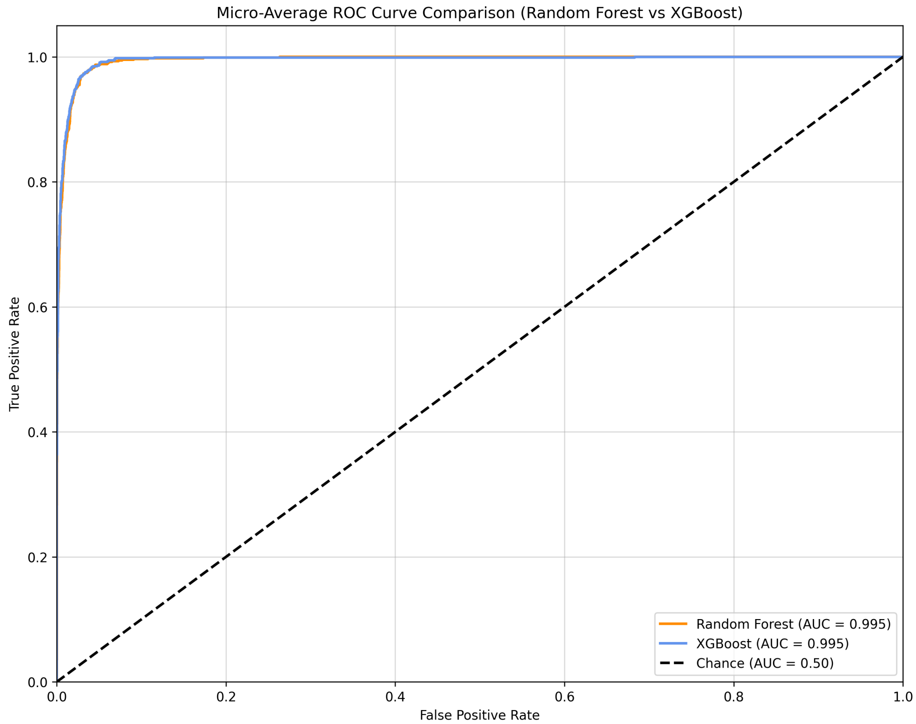
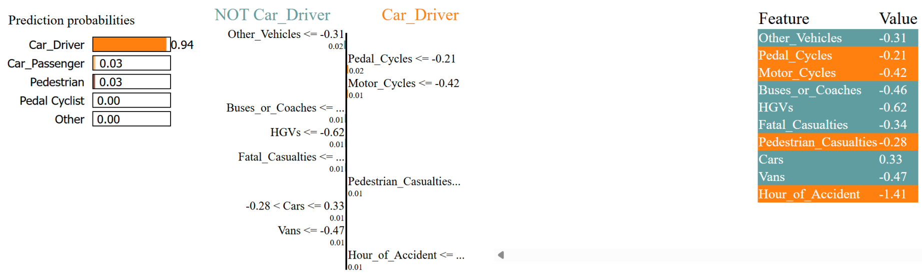


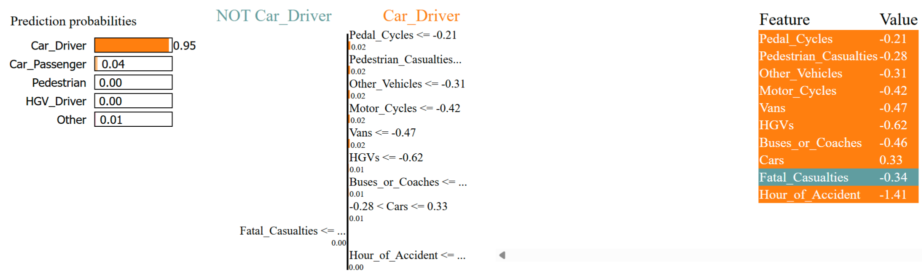
Fig. 7. Micro-Average ROC Curve for RF and XGBoost.

patterns, while LIME analysis offered local interpretability for ML models, illustrating feature contributions for individual predictions. The study underscores the potential of applying advanced NLP techniques to structured data problems via text conversion. It confirms the effectiveness of robust ML ensembles for this type of tabular prediction task. The insights gained can aid authorities in developing more targeted road safety interventions by better understanding the factors associated with different fatal casualty types.

Future Work. Future research could extend this work in several directions: *Utilizing More Recent Data.* Applying the models to current datasets to assess performance considering temporal changes [35]. *Advanced Models and Ensembles.* Exploring more sophisticated DL architectures, larger Transformers, or ensemble methods combining predictions from different model families. *Consistent Data Handling.* Re-evaluating models using identical data splits and consistently applying imbalance handling techniques for fairer comparison. *Deeper Explainability.* Applying more advanced XAI techniques (e.g., SHAP) across all model types for global and local insights [26]. Addressing these areas can further refine our understanding of fatal RTAs and contribute to developing increasingly effective global road safety strategies.



(a) Random Forest LIME explanation for instance 0 (Predicted: Car Driver).



(b) XGBoost LIME explanation for instance 0 (Predicted: Car Driver).

Fig. 8. LIME Explanations for Test Instance 0 from RF and XGBoost Models.

CRediT Authorship Contribution Statement

Umar Hasan: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – Original Draft, Visualization.
Mohammad Abdul Qayum: Writing – Review & Editing, Supervision, Project administration, Funding acquisition.

Acknowledgments. We thank Muhammad Rafsan Kabir (Research Assistant, Department of Electrical and Computer Engineering, North South University, Bangladesh) and all the anonymous peer reviewers for their sincere comments, suggestions, and criticisms.

Disclosure of Interests. The authors share organizational affiliations with several chairs of the conference committee. The authors declare no other competing interests.

Data Availability. The datasets employed for this study are publicly available. They can be accessed through https://www.data.gov.uk/dataset/73f4cd3e-92ed-4cf8-ad0b-0fe30042b626/reported-fatal-personal-injury-road-accident-and-casualty-data-gb-2006-2008.

References

1. AlMamlook, R.E., Kwayu, K.M., Alkasisbeh, M.R., Frefer, A.A.: Comparison of machine learning algorithms for predicting traffic accident severity. In: 2019 IEEE

Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT). pp. 272–276 (2019). https://doi.org/10.1109/JEEIT.2019.8717393

2. Arik, S.O., Pfister, T.: Tabnet: Attentive interpretable tabular learning. Proceedings of the AAAI Conference on Artificial Intelligence **35**(8), 6679–6687 (May 2021). https://doi.org/10.1609/aaai.v35i8.16826, https://ojs.aaai.org/index.php/AAAI/article/view/16826
3. Beshah, T., Hill, S.: Mining road traffic accident data to improve safety: Role of road-related factors on accident severity in ethiopia. In: AAAI Spring Symposium: Artificial Intelligence for Development. vol. 24, pp. 1173–1181 (2010), https://cdn.aaai.org/ocs/1173/1173-5864-1-PB.pdf
4. Bokaba, T., Doorsamy, W., Paul, B.S.: Comparative study of machine learning classifiers for modelling road traffic accidents. Applied Sciences **12**(2) (2022). https://doi.org/10.3390/app12020828, https://www.mdpi.com/2076-3417/12/2/828
5. Boo, Y., Choi, Y.: Comparison of mortality prediction models for road traffic accidents: an ensemble technique for imbalanced data. BMC public health **22**(1), 1476 (2022). https://doi.org/10.1186/s12889-022-13719-3
6. Breiman, L.: Random forests. Machine learning **45**, 5–32 (2001). https://doi.org/10.1023/A:1010933404324
7. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research **16**, 321–357 (2002). https://doi.org/10.1613/jair.953
8. Chen, M.M., Chen, M.C.: Modeling road accident severity with comparisons of logistic regression, decision tree and random forest. Information **11**(5) (2020). https://doi.org/10.3390/info11050270, https://www.mdpi.com/2078-2489/11/5/270
9. Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. p. 785–794. KDD '16, Association for Computing Machinery, New York, NY, USA (2016). https://doi.org/10.1145/2939672.2939785, https://doi.org/10.1145/2939672.2939785
10. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: Pre-training of deep bidirectional transformers for language understanding. In: Burstein, J., Doran, C., Solorio, T. (eds.) Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). pp. 4171–4186. Association for Computational Linguistics, Minneapolis, Minnesota (Jun 2019). https://doi.org/10.18653/v1/N19-1423, https://aclanthology.org/N19-1423/
11. Dogru, N., Subasi, A.: Traffic accident detection using random forest classifier. In: 2018 15th Learning and Technology Conference (L&T). pp. 40–45 (2018). https://doi.org/10.1109/LT.2018.8368509
12. Dong, S., Khattak, A., Ullah, I., Zhou, J., Hussain, A.: Predicting and analyzing road traffic injury severity using boosting-based ensemble learning models with shapley additive explanations. International Journal of Environmental Research and Public Health **19**(5) (2022). https://doi.org/10.3390/ijerph19052925, https://www.mdpi.com/1660-4601/19/5/2925
13. Hosmer Jr, D.W., Lemeshow, S., Sturdivant, R.X.: Applied logistic regression. John Wiley & Sons, 3rd edn. (2013), https://www.researchgate.net/profile/Andrew-Cucchiara/publication/261659875_Applied_Logistic_Regression/links/542c7eff0cf277d58e8c811e/Applied-Logistic-Regression.pdf

14. Kabir, M.R., Yasar, M.S.: Leveraging machine learning algorithms for improved road safety. In: 2024 IEEE International Conference on Advanced Systems and Emergent Technologies (IC_ASET). pp. 1–6 (2024). https://doi.org/10.1109/IC_ASET61847.2024.10596242
15. Khan Rifat, M.A., Kabir, A., Huq, A.: An explainable machine learning approach to traffic accident fatality prediction. Procedia Computer Science **246**, 1905–1914 (2024). https://doi.org/10.1016/j.procs.2024.09.704, https://www.sciencedirect.com/science/article/pii/S1877050924027649, 28th International Conference on Knowledge Based and Intelligent information and Engineering Systems (KES 2024)
16. Kramer, O.: K-Nearest Neighbors, pp. 13–23. Springer Berlin Heidelberg, Berlin, Heidelberg (2013). https://doi.org/10.1007/978-3-642-38652-7_2, https://doi.org/10.1007/978-3-642-38652-7_2
17. Krishnaveni, S., Hemalatha, M.: A perspective analysis of traffic accident using data mining techniques. International Journal of Computer Applications **23**(7), 40–48 (2011). https://doi.org/10.5120/2896-3788
18. Kumeda, B., Zhang, F., Zhou, F., Hussain, S., Almasri, A., Assefa, M.: Classification of road traffic accident data using machine learning algorithms. In: 2019 IEEE 11th International Conference on Communication Software and Networks (ICCSN). pp. 682–687 (2019). https://doi.org/10.1109/ICCSN.2019.8905362
19. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.: Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692 (2019). https://doi.org/10.48550/arXiv.1907.11692
20. Ludwig, M., Moreno-Martinez, A., Hölzel, N., Pebesma, E., Meyer, H.: Assessing and improving the transferability of current global spatial prediction models. Global Ecology and Biogeography **32**(3), 356–368 (2023). https://doi.org/10.1111/geb.13635, https://onlinelibrary.wiley.com/doi/abs/10.1111/geb.13635
21. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. In: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 32. Curran Associates, Inc. (2019), https://proceedings.neurips.cc/paper_files/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf
22. Popescu, M.C., Balas, V.E., Perescu-Popescu, L., Mastorakis, N.: Multilayer perceptron and neural networks. WSEAS Trans. Cir. and Sys. **8**(7), 579–588 (Jul 2009), https://dl.acm.org/doi/abs/10.5555/1639537.1639542
23. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A.: Catboost: unbiased boosting with categorical features. In: Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 31. Curran Associates, Inc. (2018), https://proceedings.neurips.cc/paper_files/paper/2018/file/14491b756b3a51daac41c24863285549-Paper.pdf
24. Rekha Sundari, M., Reddi, P., Satyanarayana Murthy, K., Sai Sowmya, D.: Fatality prediction in road accidents using neural networks. In: Mahapatra, R.P., Peddoju, S.K., Roy, S., Parwekar, P. (eds.) Proceedings of International Conference on Recent Trends in Computing. pp. 25–33. Springer Nature Singapore, Singapore (2023). https://doi.org/10.1007/978-981-19-8825-7_3

25. Ribeiro, M.T., Singh, S., Guestrin, C.: "Why should I trust you?": Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. p. 1135–1144. KDD '16, Association for Computing Machinery, New York, NY, USA (2016), https://dl.acm.org/doi/pdf/10.1145/2939672.2939778
26. Sai Ram Aditya, P., Pal, M.: Enhancing trust and interpretability of complex machine learning models using local interpretable model agnostic shap explanations. International Journal of Data Science and Analytics **18**, 457–466 (10 2023). https://doi.org/10.1007/s41060-023-00458-w
27. Sanh, V., Debut, L., Chaumond, J., Wolf, T.: Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108 (2019). https://doi.org/10.48550/arXiv.1910.01108
28. Suthaharan, S.: Decision Tree Learning, pp. 237–269. Springer US, Boston, MA (2016). https://doi.org/10.1007/978-1-4899-7641-3_10, https://doi.org/10.1007/978-1-4899-7641-3_10
29. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L.u., Polosukhin, I.: Attention is all you need. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 30. Curran Associates, Inc. (2017), https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf
30. Wen, X., Xie, Y., Jiang, L., Pu, Z., and, T.G.: Applications of machine learning methods in traffic crash severity modelling: current status and future directions. Transport Reviews **41**(6), 855–879 (2021). https://doi.org/10.1080/01441647.2021.1954108, https://doi.org/10.1080/01441647.2021.1954108
31. Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., Drame, M., Lhoest, Q., Rush, A.: Transformers: State-of-the-art natural language processing. In: Liu, Q., Schlangen, D. (eds.) Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. pp. 38–45. Association for Computational Linguistics, Online (Oct 2020). https://doi.org/10.18653/v1/2020.emnlp-demos.6, https://aclanthology.org/2020.emnlp-demos.6/
32. World Health Organization (WHO): Overview, global status report on road safety 2018. Tech. rep., World Health Organization, Geneva (2018), https://www.who.int/publications/i/item/9789241565684
33. Yan, M., Shen, Y.: Traffic accident severity prediction based on random forest. Sustainability **14**(3) (2022). https://doi.org/10.3390/su14031729, https://www.mdpi.com/2071-1050/14/3/1729
34. Yang, J., Han, S., Chen, Y.: Prediction of traffic accident severity based on random forest. Journal of Advanced Transportation **2023**(1), 7641472 (2023). https://doi.org/10.1155/2023/7641472, https://onlinelibrary.wiley.com/doi/abs/10.1155/2023/7641472
35. Zhao, Z., Chrysostomou, G., Bontcheva, K., Aletras, N.: On the impact of temporal concept drift on model explanations. In: Goldberg, Y., Kozareva, Z., Zhang, Y. (eds.) Findings of the Association for Computational Linguistics: EMNLP 2022. pp. 4039–4054. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Dec 2022). https://doi.org/10.18653/v1/2022.findings-emnlp.298, https://aclanthology.org/2022.findings-emnlp.298/