International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(Review Article)

Check for updates

Post-crash analysis and injury severity prediction in vehicle-pedestrian collisions using logistic regression and AI-based predictive analytics

Samuel Omefe *

Department of Civil Engineering, George Washington University, Washington DC, USA.

International Journal of Science and Research Archive, 2025, 16(01), 912-923

Publication history: Received on 01 June 2025; revised on 05 July 2025; accepted on 08 July 2025

Article DOI: https://doi.org/10.30574/ijsra.2025.16.1.2077

Abstract

Post-crash analysis and road safety management is critical for prediction of injury severity after a vehicle-pedestrian collision. Efficient prediction of injury severity after the occurrence of such events is necessary to guideline emergence medical care enhancement, transportation planning, as well as evidence-based safety courses of action. Traditional statistical methods, such as logistic regression, have been utilized vastly because of their transparent and interpretable results. However, recent developments in Artificial Intelligence (AI) have proposed machine learning (ML) and deep learning methods, which demonstrate better predictive capabilities, particularly on processing complicated, high-dimensional, and nonlinear crash data. This reviewed study aims to reveal how AI-based predictive analytics can improve injury severity prediction in vehicle-pedestrian crashes. It contrasts the advantages and weaknesses of both logistic regression and AI techniques in terms of their methodological aspects and examines the suitability of both approaches in different data conditions and realities of decision-making. An integrative review of recent empirical research and technical developments was performed, covering the application of ML algorithms such as decision trees, random forests, support vector machines, and gradient boosting and also deep neural networks (DNNs) and convolutional neural networks (CNNs). Model performance measures, explanation frameworks like SHAP and LIME, and real-world application scenarios in real time traffic systems were all assessed. An AI model will almost always be more accurate and flexible in its predictions than a logistic regression. The relative research proves that ensemble and deep learning models are especially useful in discovering nonlinear relationships and dealing with the class imbalance. However, logistic regression is still useful in its interpretability and less data demands. Logistic regression remains valuable due to its ease of understanding and interpretability, AI-based models offer better performance in terms of predictive performance of injury severity, their practical use requires trade-offs between accuracy and explainability and ethical concerns which create possibilities to make more nuanced and correct predictions. A hybrid modeling method based on logistic regression and explainable AI and real-time data integration is a promising direction to go in terms of both predictive accuracy and policy explainability and to improve pedestrian safety and post-crash interventions in the complex urban transportation setting.

Keywords: Logistic Regression; Artificial Intelligence; Machine Learning; Vehicle-Pedestrian Collisions; Injury Severity Prediction; Post-Crash Analysis

1. Introduction

Vehicle-pedestrian collisions are a major contributor to traffic-related injuries and fatalities; pedestrian safety has become critical due to increasing complexity of urban mobility systems. As people have begun to own more vehicles and the population densities in urban centers have increased, the problem of pedestrian-vehicle crashes has become a constant threat to public safety in both high and low-income nations [1]. The incidents usually contribute to serious

* Corresponding author: Samuel Omefe

injuries or even deaths and create significant burden on the health care systems revealing the systemic inefficiency in the road safety policy and emergency situation response system [2].

Traditional attempts to forecast the severity of injuries sustained by pedestrians after collision have been based on the use of statistical approaches with the logistic regression being the most prominent [3]. Although this method has offered a basis of risk factor comprehension, its linear assumptions along with the ability to represent only simple, single dimension associations among crash variables have limited its applicability [4]. With an increasing digitalization of urban spaces, the idea of using Artificial Intelligence (AI) and machine learning (ML) to improve the models of injury severity prediction and make safety interventions more proactive is gaining popularity.

The aim of the review is the examination of the shift towards advanced AI-based predictive modelling in favor of traditional post-crash analytic methods. Through the review of empirical research, practical implementations, and the developing hybrid approaches, we will evaluate the existing strengths and weaknesses of those methods in predicting the injury outcomes in the vehicle-pedestrian crashes. We also consider the ways in which these models are being incorporated into larger traffic safety systems and their potential to guide policy, optimize emergency care, and aid in the creation of autonomous transportation systems.

2. Relevance of Injury Severity Prediction in Pedestrian Safety

Correct estimation of the severity of the injury after pedestrian-vehicle crashes is crucial to various areas of public health and traffic safety administration. The main component of post-crash response systems is the capacity to rapidly assess the likelihood of fatal / life-threatening injuries to enable emergency medical services to prioritize resources, activate trauma response protocols, and make decisions in time about transport and treatment facilities [5].

In addition to direct clinical uses, models of injury severity have been used to influence the policies and tactics of transportation planners and policymakers [6]. By revealing trends in crash conditions such as the time of day, type of vehicle, road surface type or pedestrian characteristics such models can inform the delivery of evidence-based infrastructure interventions [7]. These involve implementation of speed-calming procedures in high-risk areas, improved lighting of poorly visible areas and safe pedestrian crossing enforcement [8].

Insurance companies and laws also benefit from predictive analysis of injury severity [9]. Proper modeling will help in expediting the claims process, fair determination of liability, as well as policymaking on compensation threshold. According to the development of autonomous vehicles, injury prediction models are used to assist in the calibration of the decision-making algorithms, so that in different working conditions, the priority is to keep pedestrians safe [10]. Since the utility of such models is so broad, improving their predictive value is not just a technical goal, but rather a multidimensional necessity, traversing health care, governance, city form, and social justice.

3. Foundation and Limits of Logistic Regression in Injury Severity Analysis

In the transportation research area, logistic regression has been a long time the statistical workhorse of injury severity modeling. Because it is a classification algorithm that can approximate the likelihood of discrete variables, e.g., “minor,” “severe,” or “fatal” injuries, it has seen wide use because it is easily interpretable and fits well with categorical data [11-12].

One of the initial fields of the use of the logistic regression was the evaluation of pedestrian injuries, to determine the correlations between the injury results and vehicles speed, pedestrian age, crossing patterns, and environmental factors [13]. Indicatively, studies based on the Fatality Analysis Reporting System (FARS) in the United States consistently demonstrated that elderly pedestrians and greater vehicle speeds are strong indicators of death injuries [14]. Equally, the road lighting conditions and driver distraction have also been stated as common explanatory variables in the multinomial logistic models [15].

Logistic regression is limited by the methodological shortcomings. It presumes that the independent variables have a linear association with the log-odds of the dependent variable, which is frequently not the situation in the actual crash data which frequently exhibit complicated interrelationships and non-linear impacts [16]. Logistic models are vulnerable to multicollinearity and need cautious variable choosing, and they might not perform well with high-dimensional data, especially the data collected in sensor-dense smart environments or high-precision vehicle telemetry systems [17].

Logistic regression is advantageous because of its interpretative transparent nature. Traffic safety professionals and policymakers commonly prefer models in which each variable effect is clearly defined and may be intuitively translated into a policy or regulatory measure [18]. Because of this, logistic regression has remained a staple of post-crash analysis, despite more sophisticated machine learning methods coming to the fore.

4. AI-Based Predictive Analytics in Injury Severity Prediction

Using AI-based predictive analytics in post-crash injury modeling is a transport safety research paradigm shift. Such models, especially those based on machine learning and deep learning frameworks, have become notable because of their capability to estimate nonlinear relationships in high dimensional crash data, which in most cases cannot be achieved by the conventional statistical methods [19].

4.1. Methods of Machine Learning

Machine learning (ML) is a collection of algorithms that can learn, without being explicitly programmed, based on data patterns. Decision trees, random forests, support vector machines (SVM) and gradient boosting (XGBoost) are among the most popular algorithms used in injury severity prediction [20]. Such models have shown to have great predictive power in managing interaction of variables, class imbalance and non-linear relationships.

Decision trees offer a natural branching format in which the data is divided into different risk groups by the thresholds of features [21]. Random forests built upon this idea by ensemble-averaging the predictions of many decision trees, which can decrease the variance and enhance generalization [22]. SVMs on the other hand build hyperplanes in a high dimensional space to ensure noise resistant classification of the outcome of injuries [23]. The gradient boosting models are known to iteratively rectify the prediction errors and thus they are very flexible to intricate injury severity profiles [24]. These models have been very useful especially in huge crash databases where there are many features such as environmental factors, vehicle characteristics among many others whose interactions are complex. Compared to logistic regression, ML models can enable more subtle risk stratification by finding obscure combinations of variables that are associated with risk of injury.

4.2. Neural Networks and Deep Learning

The future of predictive analytics in traffic safety is deep learning models, namely deep neural networks (DNNs) and convolutional neural networks (CNNs). These are types of models which automatically learn hierarchical feature Representations (features) from unprocessed input data, whether that be time-series telemetry, crash scene imagery or even driver and pedestrian behavior videos [25-26].

DNNs, having appeared hidden layers, are capable of discovering high-level, complex abstractions in injury severity data. CNNs excel particularly with spatial and visual data, which is why they are well-suited to interpret images captured by surveillance units or dashcams [27]. Trained with large, labeled crash datasets, these models may predict injury severity in a degree of detail, and with a sense of context, that was previously not possible. They are complex and need many computational resources and access to high-resolution data that is well labeled. Moreover, such models are not easily interpretable without further explanatory measures, thus they are not favorable in situations where decision transparency is needed [28].

4.3. Feature Engineering / Selection

One of the fundamental elements of AI modeling is feature engineering, as it defines the quality and relevance of inputs to be utilized in prediction [29]. In models such as logistic regression, feature selection, usually through domain knowledge, is of great importance. However, AI methods have the ability to automate feature extraction and transformation [30].

Dimensionality reduction can be performed using techniques like Principal Component Analysis (PCA), autoencoders and recursive feature elimination to alleviate the dimensionality curse and to make the models efficient [31]. These techniques not only improve performance, but also help in alleviating the problem of overfitting that is typical of high dimensional crash data [32]. Tree based algorithms intrinsically provide feature importance ranking which allows one to identify the most predictive variables without significant preprocessing [33]. Such a feature allows adaptive modeling, which changes when new data or sensor readings are available.

4.4. Model Assessment

Assessing the performance of predictive models is important to guarantee reliability particularly in safety-critical tasks such as injury severity prediction. Whereas the common measures of assessment of the logistic regression models are accuracy and p-values, AI models need a wider range of performance metrics, as they are probabilistic and multi-class [34].

The important evaluation criteria include accuracy, which is the percentage of correct predictions among all classes. Precision, recall, and F1-score, which evaluate the model capability to accurately predict certain classes of injuries, especially relevant in imbalanced datasets. Area Under the Receiver Operating Characteristic Curve (AUC-ROC), a common metric to compare the performance of classifiers, is based on the trade-off between sensitivity and specificity and needs to be appropriately validated in order to confirm the generalizability and clinical relevance of AI-based injury severity models [35-36].

4.5. AI Model Explainability and Trust

However, in spite of the strengths associated with predictive models, AI models have been accused of lacking explainability. Many machine learning and deep learning models do not give much explanation on the decision-making process, contrary to logistic regression which has coefficients that can be interpreted. In response thereto, explainable AI (XAI) guided the introduction of interpretability frameworks like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) [37]. These tools examine the contribution of every feature to the individual predictions and improve the transparency and accountability of the models [38]. SHAP values can in practice measure the role of the vehicle speed or age of pedestrian on a particular injury outcome, bringing the reasoning of the model closer to policymakers, clinicians and legal experts. Incorporation of such explainability methods will be crucial to the development of trust in AI systems, as the latter become incorporated into emergency response dispatch systems, insurance systems, and self-driving car algorithms [39].

5. Integrated Perspective on Injury Severity Prediction

Injury severity prediction is vital for improving emergency response and safety measures in vehicle-pedestrian collisions. While traditional methods like logistic regression offer clear insights, they often struggle with complex data patterns. Advances in AI and machine learning provide more accurate and adaptable predictive models, capable of handling diverse and high-dimensional data.

5.1. The Emergence of AI-Based Predictive Analytics in Post-Crash Injury Modeling

The introduction of the Artificial Intelligence (AI) into traffic safety research has reinvented the methodological infrastructure of injury severity prediction. Although logistic regression gives structured and interpretable models, AI-based predictive analytics capabilities to identify nonlinear relationships, trained on large amounts of heterogeneous data, and evolve with shifting patterns in crash surroundings [40]. Their usefulness can be especially in an urban environment, where the pedestrian-vehicle interaction depends on many moving, interrelated factors.

Machine learning algorithms including decision trees, random forests, support vector machines (SVM), gradient boosting approaches, and deep neural networks AI models have been progressively used in injury prediction tasks [41]. They are appealing because they can take structured and unstructured data, such as images, sensor data, geospatial data, traffic patterns, and behavioral signals of smart city infrastructure [42]. As an example, convolutional neural networks (CNNs) have been applied to analyze dashcam footage and surveillance video to predict the likelihood of injury depending on the physical mechanics of the crash.

This is one of the essential advantages of AI in the field because, unlike other methods, it may reveal the existence of intricate relations among the crash variables, without any a priori assumption about the form of the functional relationships among them [43]. In contrast to logistic regression, where feature engineering and hypothesis-driven variable selection is required manually, machine learning methods have the potential to capture latent patterns via automated methods of feature importance ranking and dimensionality reduction [44]. This enables the utilization of high-dimensional data, e.g. the data gathered by connected vehicles, roadside sensors, and wearable technologies.

The AI models have demonstrated an outstanding performance with predictive accuracy. Ensemble learning methods such as random forests and XGBoost have regularly beaten the traditional models in the comparative studies in the classification of injury outcomes, particularly in the imbalanced classes a usual situation in the crash severity modeling where non-fatal cases are overwhelmingly more than the fatalities [19]. The use of deep learning techniques further improves performance and learns hierarchical representations directing on raw inputs which enables the model to generalize fine-grained risk profiles on the basis of spatial, temporal, and behavioral inputs [45]. Even though these developments are taking place, the usage of AI in predicting injury severity is not flawless. The absence of transparency that is linked to most AI algorithms is one of the most urgent concerns [38]. In contrast to logistic regression, where the coefficient estimates and p-values are provided. Entities models based on AI that may work as so-called black boxes, they provide correct predictions but do not offer insight into the reasoning behind these predictions [46]. Such opaqueness has the potential to impede their usage in policy and legal contexts where such traceability and accountability are crucial.

As a reaction to this shortcoming, explainable AI (XAI) has become popular. SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) are tools that allow a researcher to interpret model predictions by attributing importance scores to each predictor feature, per prediction [38]. Such tools are becoming more applicable to make AI models more palatable to non-technical traffic safety and populace health stakeholders. Another interesting advancement is the implementation of prediction systems using AI into reality in urban mobility. Smart cities have already started using predictive analytics in advanced traffic management systems that can predict high-risk areas and send emergency services before they are needed [47]. A portion of autonomous vehicle platforms are also integrating AI-based severity estimation models to guide evasive maneuvers, braking interventions, and post-collision response plan.

AI-based predictive analytics are not merely replacing traditional statistical methods but rather that, they are opening up new frontiers on what can be measured, predicted, and acted upon when it comes to the prevention of pedestrian injury. Their further development, particularly in combination with the attempts to make them more transparent and generalizable has a great potential to make road safety interventions more precise and fairer.

5.2. Relative Analysis of Logistic Regression and AI-Based Models

There is an emerging literature comparing logistic regression models and Artificial Intelligence (AI)-based predictive models with regard to estimating injury severity after vehicle-pedestrian crashes [48]. These comparative studies indicate the existence of a central trade-off between interpretability and predictive performance and also show important differences in the manner each modeling approach treats data complexity, variable interaction, and class imbalance [49].

Logistic regression has been long considered as the gold standard of injury severity modelling because of its parsimony, interpretability and well-developed statistical foundation [18]. It has direct connection with the direction and strength of association between predictors and the outcome since its coefficients can be directly interpreted, which makes it especially useful in explanatory modeling. In other words, old pedestrians or fast cars may be directly applied to community safety campaigns or legal codes when a logistic regression model finds them to be strongly related to fatal outcomes. The ability to output confidence intervals and statistical significance of each predictor is also a useful option to have, especially in a policy-based setting where decisions must be substantially justified [16].

Logistic regression, when calculated based on predictive accuracy alone, many times underperforms AI-based models. This has been confirmed by many publications based on real-world crash data that ensemble machine learning algorithms, including random forests and gradient-boosting machines, invariably produce superior classification statistics. Such models can learn non-linear relationships and high-order interactions among variables features that logistic regression does not have unless those complexities are pre-programmed into the model by hand. To give some examples, in a study that involved the U.S. Fatality Analysis Reporting System (FARS), a random forest model had an area under the curve (AUC) of 0.80 in pedestrian injury severity prediction, which is significantly higher than logistic regression [50].

The capability of the AI models to handle big and multi-modal data is one of the factors that allow them to thrive. They are able to consume not just tabular data like vehicle type, weather conditions and time of day but also visual data in the form of surveillance video or spatial data in the form of GPS and road network data [51]. Research Deep Neural Networks, most prominently convolutional neural networks (CNNs), have been used in research to process imagery and spatial arrangements at crash sites, which shows further promise in improving injury severity predictions over and above what is provided by tabular data alone [25].

In spite of these benefits, AI models are experiencing a credibility problem in most practical applications because of their low explainability. Many algorithms are opaque, which may be an obstacle to safety-related fields where responsibility, explainability, and interpretability are priorities. Policy makers and lawyers tend to feel uneasy about implementing suggestions made by models which do not offer easily comprehendible logic [38]. Although partial solutions such as SHAP and LIME allow a certain degree of clarity and policy relevance that traditional statistical coefficients provided, they do not necessarily replicate it.

The logistic regression models tend to be more stable when the dataset is small or when there are fewer variables, and therefore it is more feasible to use in jurisdictions or institutions that do not have extensive crash database or technology base [52]. Conversely, AI models generally need large well-annotated datasets to train well, and might perform much worse when noisy, when values are missing, or when classes are imbalanced unless measures are taken to counteract these effects, for example, by data augmentation or resampling.

The second emerging approach is the creation of hybrid models, which integrate the explainability of logistic regression and the performance benefits of machine learning. Logistic regression can be applied in such models to provide baseline relationships, with machine learning layers excess, non-linear variance. Others have used model-stacking methods, in which the logistic regression model results are used as features to a machine learning meta-model, improving prediction performance and interpretability [53].

Specific goals of the analysis, the type of available data, and stakeholders are to influence the decision between logistic regression and AI-based approaches. Logistic regression could still be the tool of choice, though, when the major objective is explanatory knowledge to guide policymaking or legal adjudication. Nevertheless, when predictive accuracy is of particular importance, particularly in real-time applications or high-dimensional data AI-based analytics have better capabilities. With the richness of data sources and the maturation of AI interpretability tools, convergence of the two methods seems not just possible, but very attractive.

5.3. Comparative Studies on Predictive Performance

An increasing number of empirical studies have compared the traditional logistic regression to AI-based predictive models as applied to crash injury severity. Even though logistic regression is still appreciated due to its interpretability and statistical clarity, machine learning algorithms repeatedly outperform it in terms of accuracy and flexibility when confronted with different data.

A study conducted using a worldwide meta-analysis of 56 crash-injury data arrays found that the ensemble techniques especially the Random Forest worked better than the logistic regression in about 70% of the cases [41]. Another study utilized accident data from the N-5 Highway of Pakistan to compare LightGBM, CatBoost, XGBoost, and logistic regression. The model that provided the best accuracy was LightGBM with 73.6% accuracy, 70.8% F1-score, and the AUC of 0.71 [54]. A study used a Random Forest classifier on a metropolitan accident dataset in the U.S and achieved an accuracy of more than 80%, precision of 97.1%, recall of 79.2%, F1-score of 87.3%, and an AUC of 0.80. Best predictors that were also discovered in the study included wind speed, pressure, humidity, visibility, clear weather, and cloud cover [50].

Table 1 Study Dataset Methods Compared Best Performing Model

Study	Dataset / Context	Methods Compared	Best Model and Performance
[41]	Global meta-analysis	LR, Random Forest, SVM, Decision Tree	Random Forest – best performer in ≈ 70% of cases
[54]	Pakistan N-5 Highway	LightGBM, CatBoost, XGBoost, LR	LightGBM – 73.6% accuracy, F1=70.8%, AUC = 0.71
[50]	U.S. metropolitan accidents	Random Forest	RF – Accuracy >80%, Precision 97.1%, Recall 79.2%, F1 = 87.3%, AUC = 0.80

These studies uniformly advocate the excellence of ensemble and boosting machine learning model in predicting injury severity, but they do not refute the interpretive advantages of logistic regression, especially where stakeholder explainability and resource limitation are paramount.

6. Challenges and Ethical Considerations in Predictive Injury Modelling

There are methodological and ethical issues with integrating predictive analytics into the post-crash injury severity assessments process. This is especially acute in the case of using traditional statistical models to AI-based systems, which are more data-hungry, complicated, and obscure in decision-making mechanics. In this regard, specific consideration needs to be given to the quality of data, its interpretability, fairness, privacy, or unintended consequences that can emerge in practical applications.

6.1. Quality and Availability of Data

The problem of data quality is one of the most insistent in modeling the severity of injuries, no matter what methodological approach is used. A common problem with many national and regional crash databases is varying degrees of inconsistency in reporting, missing values, underreporting of non-fatal injuries, and a lack of information relating to many contextual variables, for example, pedestrian behavior, pre-crash conditions [55]. Such shortcomings may undermine the effectiveness of logistic regression models and the generalization of AI systems, which in many cases need large clean dataset to train and infer efficiently.

Besides that, datasets including sensor readings, dashcams videos, or telematics data are frequently owned and not available to researchers because of either commercial rights or data protection regulations [56]. This limits the scale of AI-based methods and does not allow cross-jurisdictional comparison of models. Accuracy and fairness in predictive modeling cannot be ensured without regular access to high-quality standardized data.

6.2. Transparency and Interpretability of Model

The main ethical concern of a predictive model is the understandability of the tools used. Logistic regression provides a simple, easily interpreted model in which the meaning of each coefficient is clear, which is appropriate to be used in policy and discussion with the population. Conversely, most machine learning and AI models especially deep neural networks are deemed as black boxes, and it is not easy to trace the contribution of input features to predictions [46].

This opportunity lacks transparency that is of concern when applying models to high-stakes decisions, like assigning fault in a legal proceeding or responding to emergencies. When a model incorrectly labels a life-threatening injury as minor, or a court is requested to depend on an inexplicable AI forecast in a liability designation, the ramifications may be tremendous and ethically dubious. Explainable AI advances recently achieved some progress on these issues, although the current tools still cannot offer the complete accountability that is required in most cases of public safety.

6.3. Discrimination and Equality

Data and model bias is a serious ethical issue in AI-based injury severity prediction. Any difference in the crash reporting that depends on the geographical area, socioeconomic status, or any demographic factors may produce biased training data, and consequently, it will promote the same inequalities. As an example, consider a case where the collision data used to train models do not represent collisions with marginalized communities well [57]. In this case, models can be poorly calibrated when predicting the severity of injuries on marginalized community members, resulting in biased emergency interventions or inefficiently allocated safety resources.

Besides, when demographic variables are not provided as inputs to a model to explicitly discriminate, proxy variables can meanwhile embed the same biases [58]. The challenge of fairness in predictive modeling could only be solved by not only technical means, like algorithmic debiasing and fairness-aware learning, but also through permanent ethical monitoring, consultation with stakeholders, and explanations of model development and use.

6.4. The issue of Privacy and Surveillance

The growth of real-time data dependence on smart infrastructure, connected automobiles, and wearable items emerges with urgent issues of privacy and monitoring. The use of continuous geolocation tracking, behavioral monitoring, or biometric data in injury severity modeling brings into question the manner in which this data is generated, accessed, and what other applications it could be put when not applied in the safety field.

In areas where data protection laws are weaker, a danger lies in the possibility of surveillance-based predictive analytics being re-purposed to non-consensual monitoring, or discriminatory application. Moreover, predictive models utilized in an automated decision-making system, e.g. autonomous cars or smart city platforms, can encroach on individual freedoms and circumvent the due process without adequate regulation [59].

By centralising and exposing less data, some of these risks can be alleviated through the use of privacy-preserving AI methods, e.g. federated learning or differential privacy. Such solutions, however, must be supported by law, code of conduct, and institutional countermeasures that ensure that the rights of an individual are not violated but that which serves the greater good.

6.5. Real-world Implementation and Organization Preparedness

The effective implementation of predictive models (AI) in a transportation safety system is also dependent on the wider institutional preparedness. Most cities and government agencies do not have the technical systems, talented workforce, or regulatory systems to implement and manage AI technologies in a responsible manner [60]. In addition, interoperability, standardized protocols and strong accountability measures are necessary in integration with emergency response systems, law enforcement and healthcare services.

7. Conclusion

With the predicted increase in population density in cities and complexity of their mobility systems, pedestrian safety is becoming an immediate concern of researchers, public officials, and transportation engineers. Effective estimation of the seriousness of injuries caused by vehicle-pedestrian impacts is important not merely to enhance short-term emergency hold but also to guide long-term road safety, urban planning, and regulatory policy-making.

The review has covered the history of injury severity modelling, starting with the seminal place of logistic regression and heading to the more flexible and customizable world of AI-driven predictive analytics. Although logistic regression has given decades of useful information about the association accompanying crash-related variables and injury outcome, its drawbacks, especially its inability to deal with complex, non-linear, and high-dimensional data have paved the way to more elaborate modelling strategies. Ensemble models, machine learning algorithms and deep learning architectures have proven to be more accurate in their predictions which present new and powerful tools to aid in post-crash analysis and decision making.

But this transition is not trouble-free. The growing sophistication of AI models is creating severe issues concerning explainability, responsibility, and ethical use in particularly high-stakes situations, including emergency response and liability assessment. Data quality, surveillance, bias and digital exclusion raise concerns on the importance of having rigorous governance structures and inclusive model design.

The way ahead is not replacement of the old ways en masse, but careful combination of logistic regression and AI. A fascinated solution is presented by the hybrid models that integrate the statistical explainability with the algorithmic flexibility, especially when augmented by the explainer AI frameworks and interoperable, standardized data systems. Moving forward, the future of injury severity prediction will rely on the future interdisciplinary collaboration. The collaboration of data scientists, clinicians, traffic engineers, legal experts and community stakeholders is needed to assure that predictive tools are not merely technically competent, but also morally acceptable and responsible to the society. Post-crash predictive analytics has the potential to be transformative in terms of injury severity reduction, efficient emergency response, and even lives saved by filling the gap between interpretability and innovation.

Recommendations

There is an urgent need to encourage the emergence of hybrid analytical models that combine the explainability of logistic regression and the predictive performance of machine learning and AI methods. Such models may play an explanatory role as well as an operational role of closing the divide between policy relevance and technical performance. The research directions must work on optimizing model architectures beyond accuracy to transparency and flexibility to heterogeneous data environments.

The essential investment needs to be made in high-quality standardized interoperable data infrastructure. Transportation agencies, governments and healthcare institutions ought to liaise in the formation of consolidated data collection measures, which would cut across crash reporting, EMS response, hospital records and environmental influences. Reproducibility would be increased, comparative analysis would become possible, and model development would be speeded by open-access repositories where the definitions of variables and standards of annotation were consistent.

Explainability and fairness AI injury predictive models need to be deployed together with explainability and fairness mechanisms. The stakeholders ought to embrace explainable AI systems and perform frequent audits to identify and neutralize biases in algorithms, particularly those that harm susceptible groups. The development of ethical guidelines should be imposed so that AIs for their decisions could be transparent, accountable, and contestable.

The application of real-time predictive analytics is something that needs to be piloted in smart city settings and specifically in urban corridors that are at a higher risk of pedestrian-related accidents due to the exposed nature of the pedestrians. Such pilots must be judged not solely on technical performance, but also on popular acceptability, interoperability of systems and fairness of access. Emergency response systems should be prepared to handle these predictions so that they become a normal part of the dispatch procedures and triaging of treatments.

There is a need to build capacity and involve stakeholders to integrate predictive analytics in road safety ecosystems in a successful way. Policy makers, traffic law enforcement, first responders, and city planners need to be informed about the capability and shortcomings of predictive models. This will enable them to trust and be relevant in the real-world application since they will be actively involved in the model design, evaluation, and deployment processes. Inclusive innovation must be prioritized with urgent need to customize predictive tools so they could be applicable in resource-limited environments. Also, many areas where pedestrian injuries are the most severe do not have the digital infrastructure to use AI-based solutions. Funding and research development ought to endorse lighter, more context-sensitive models capable of functioning with little hardware, sparse information, and minimal bandwidth.

References

[1] Maurer O. The Case of Santiago de Chile: Pedestrian Deaths, Neo-Liberal Urban Design, and Insufficient Traffic Policy Reform.

[2] Gelderblom AJ. Analysis of the relationship between the severity of road traffic crashes and the human factors involved: N4 Toll route case study (Doctoral dissertation, Stellenbosch: Stellenbosch University).

[3] Tao W, Aghaabbasi M, Ali M, Almaliki AH, Zainol R, Almaliki AA, Hussein EE. An advanced machine learning approach to predicting pedestrian fatality caused by road crashes: A step toward sustainable pedestrian safety. Sustainability. 2022 Feb 20;14(4):2436.

[4] Islam MK, Reza I, Gazder U, Akter R, Arifuzzaman M, Rahman MM. Predicting road crash severity using classifier models and crash hotspots. Applied Sciences. 2022 Nov 9;12(22):11354.

[5] Valente JT. From Crash to Care: A Road Towards Improved Safety and Efficiency of Emergency Medical Response.

[6] Forjuoh SN, Guohua LI. A review of successful transport and home injury interventions to guide developing countries. Social science and medicine. 1996 Dec 1;43(11):1551-60.

[7] Goel R, Tiwari G, Varghese M, Bhalla K, Agrawal G, Saini G, Jha A, John D, Saran A, White H, Mohan D. Effectiveness of road safety interventions: An evidence and gap map. Campbell systematic reviews. 2024 Mar;20(1):e1367.

[8] World Health Organization. Pedestrian safety: a road safety manual for decision-makers and practitioners. World Health Organization; 2023 May 3.

[9] Zhang W, Shi J, Wang X, Wynn H. AI-powered decision-making in facilitating insurance claim dispute resolution. Annals of Operations Research. 2023 Oct 23:1-30.

[10] Tang X, Yang K, Wang H, Wu J, Qin Y, Yu W, Cao D. Prediction-uncertainty-aware decision-making for autonomous vehicles. IEEE Transactions on Intelligent Vehicles. 2022 Jul 5;7(4):849-62.

[11] Bhuiyan H, Ara J, Hasib KM, Sourav MI, Karim FB, Sik-Lanyi C, Governatori G, Rakotonirainy A, Yasmin S. Crash severity analysis and risk factors identification based on an alternate data source: a case study of developing country. Scientific reports. 2022 Dec 8;12(1):21243.

[12] Hua Y, Stead TS, George A, Ganti L. Clinical risk prediction with logistic regression: Best practices, validation techniques, and applications in medical research. Academic Medicine and Surgery. 2025 Mar 6.

[13] Wang K, Zhang W, Jin L, Feng Z, Zhu D, Cong H, Yu H. Diagnostic analysis of environmental factors affecting the severity of traffic crashes: From the perspective of pedestrian–vehicle and vehicle–vehicle collisions. Traffic injury prevention. 2022 Jan 2;23(1):17-22.

[14] Hossain A, Monica NJ, Ahmad MT, Choya FN. INVESTIGATING OLDER PEDESTRIAN FATAL CRASH CHARACTERISTICS IN THE UNITED STATES.

[15] Alsaleh R, Walia K, Moshiri G, Alsaleh YT. Traffic Collision Severity Modeling Using Multi-Level Multinomial Logistic Regression Model. Applied Sciences. 2025 Jan 16;15(2):838.

[16] Dey D, Haque MS, Islam MM, Aishi UI, Shammy SS, Mayen MS, Noor ST, Uddin MJ. The proper application of logistic regression model in complex survey data: a systematic review. BMC Medical Research Methodology. 2025 Jan 22;25(1):15.

[17] Ma C, Peng Y, Wu L, Guo X, Wang X, Kong X. Application of machine learning techniques to predict the occurrence of distraction-affected crashes with phone-use data. Transportation research record. 2022 Feb;2676(2):692-705.

[18] Mansour M, Alsulamy S, Alghamdi S, Qureshi MR. Assessing the risk of university-student driving to enhance safe and sustainable traffic management: a logistic regression and analytic hierarchy approach. Tehnički vjesnik. 2022 Apr 15;29(2):692-701.

[19] Jamal A, Zahid M, Tauhidur Rahman M, Al-Ahmadi HM, Almoshaogeh M, Farooq D, Ahmad M. Injury severity prediction of traffic crashes with ensemble machine learning techniques: A comparative study. International journal of injury control and safety promotion. 2021 Oct 2;28(4):408-27.

[20] Singh S, Ramkumar KR, Kukkar A. Machine learning techniques and implementation of different ML algorithms. In2021 2nd Global Conference for Advancement in Technology (GCAT) 2021 Oct 1 (pp. 1-6). IEEE.

[21] Mienye ID, Jere N. A survey of decision trees: Concepts, algorithms, and applications. IEEE access. 2024 Jun 19.

[22] Rane N, Choudhary SP, Rane J. Ensemble deep learning and machine learning: applications, opportunities, challenges, and future directions. Studies in Medical and Health Sciences. 2024 Jul 4;1(2):18-41.

[23] Li W. A Big Data Approach to Forecast Injuries in Professional Sports Using Support Vector Machine. Mobile Networks and Applications. 2024 Sep 3:1-7.

[24] Tamim Kashifi M, Ahmad I. Efficient histogram-based gradient boosting approach for accident severity prediction with multisource data. Transportation research record. 2022 Jun;2676(6):236-58.

[25] Razi A, Chen X, Li H, Wang H, Russo B, Chen Y, Yu H. Deep learning serves traffic safety analysis: A forward-looking review. IET Intelligent Transport Systems. 2023 Jan;17(1):22-71.

[26] Shi L. Enhanced Feature Representation in Multi-Modal Learning for Driving Safety Assessment.

[27] Shadab Siddiqui MA, Rabbi MS, Islam MJ, Ahmed RU. Comparison of Different Controller Architectures for Autonomous Driving and Recommendations for Robust and Safe Implementations. Journal of Advanced Transportation. 2025;2025(1):9995539.

[28] Li, Xuhong, et al. "Interpretable deep learning: Interpretation, interpretability, trustworthiness, and beyond." Knowledge and Information Systems 64.12 (2022): 3197-3234.

[29] Zhong S, Zhang K, Bagheri M, Burken JG, Gu A, Li B, Ma X, Marrone BL, Ren ZJ, Schrier J, Shi W. Machine learning: new ideas and tools in environmental science and engineering. Environmental science and technology. 2021 Aug 17;55(19):12741-54.

[30] Sufi FK, AlSulami M. AI-feature: A software for automatic extraction of features attributes from social medial texts using AI services. International Transaction Journal of Engineering, Management, and Applied Sciences and Technologies. 2022 Mar;13(4):1-8.

[31] Tufail S, Iqbal H, Tariq M, Sarwat A. A Hybrid Machine Learning-Based Framework for Data Injection Attack Detection in Smart Grids Using PCA and Stacked Autoencoders. IEEE Access. 2025 Feb 19.

[32] Shengwen G, Osman MH. Abnormal Data Detection Model Based on Autoencoder and Random Forest Algorithm: Camera Sensor Data in Autonomous Driving Systems. International Journal of Advanced Computer Science and Applications. 2025 Mar 1;16(3).

[33] Vieira-Manzanera E, Patricio MA, Berlanga A, Molina JM. Analyzing feature importance with neural-network-derived trees. Neural Computing and Applications. 2025 Feb;37(5):3419-33.

[34] Obioha Val O, Lawal T, Olaniyi OO, Gbadebo MO, Olisa AO. Investigating the feasibility and risks of leveraging Artificial Intelligence and open source intelligence to manage predictive cyber threat models. Temitope and Olaniyi, Oluwaseun Oladeji and Gbadebo, Michael Olayinka and Olisa, Anthony Obulor, Investigating the Feasibility and Risks of Leveraging Artificial Intelligence and Open Source Intelligence to Manage Predictive Cyber Threat Models (January 23, 2025). 2025 Jan 23.

[35] Naidu G, Zuva T, Sibanda EM. A review of evaluation metrics in machine learning algorithms. InComputer science on-line conference 2023 Apr 3 (pp. 15-25). Cham: Springer International Publishing.

[36] Maddireddy BR, Maddireddy BR. Automating Malware Detection: A Study on the Efficacy of AI-Driven Solutions. Journal Environmental Sciences And Technology. 2023;2(2):111-24.

[37] Ahmad T, Katari P, Pamidi Venkata AK, Ravi C, Shaik M. Explainable AI: Interpreting Deep Learning Models for Decision Support. Advances in Deep Learning Techniques. 2024;4(1):80-108.

[38] Bhattacharya A. Applied Machine Learning Explainability Techniques: Make ML models explainable and trustworthy for practical applications using LIME, SHAP, and more. Packt Publishing Ltd; 2022 Jul 29.

[39] Kashifi MT. Investigating two-wheelers risk factors for severe crashes using an interpretable machine learning approach and SHAP analysis. IATSS research. 2023 Oct 1;47(3):357-71.

[40] Yang S. Developing Real-Time Crash Prediction System using AI-Based Methods on Interstate Incorporating Express Lanes.

[41] Santos K, Dias JP, Amado C. A literature review of machine learning algorithms for crash injury severity prediction. Journal of safety research. 2022 Feb 1;80:254-69.

[42] Leckey C, Van Dyk N, Doherty C, Lawlor A, Delahunt E. Machine learning approaches to injury risk prediction in sport: a scoping review with evidence synthesis. British Journal of Sports Medicine. 2025 Apr 1;59(7):491-500.

[43] Kobre DW. Analyzing Crash Factors Using Artificial Intelligence for Curved Rural Roads. Aspen University; 2022.

[44] Su Y, Wang X, Ye Y, Xie Y, Xu Y, Jiang Y, Wang C. Automation and machine learning augmented by large language models in a catalysis study. Chemical Science. 2024;15(31):12200-33.

[45] He T, Jin X, Zou Y. Deep learning-based action recognition for joining and welding processes of dissimilar materials. Frontiers in Materials. 2025 Mar 20;12:1560419.

[46] Marcus E, Teuwen J. Artificial Intelligence and explanation: How, why, and when to explain black boxes. European Journal of Radiology. 2024 Feb 23:111393.

[47] Mohanty A, Mohapatra AG, Mohanty SK. Real-time traffic monitoring with AI in smart cities. InInternet of Vehicles and Computer Vision Solutions for Smart City Transformations 2025 Feb 20 (pp. 135-165). Cham: Springer Nature Switzerland.

[48] Mohanty M, Panda R, Gandupalli SR, Sonowal D, Muskan M, Chakraborty R, Dangeti MR. Development of crash prediction models by assessing the role of perpetrators and victims: a comparison of ANN and logistic model using historical crash data. International journal of injury control and safety promotion. 2023 Apr 3;30(2):155-71.

[49] Kruschel S, Hambauer N, Weinzierl S, Zilker S, Kraus M, Zschech P. Challenging the Performance-Interpretability Trade-off: An Evaluation of Interpretable Machine Learning Models. Business and Information Systems Engineering. 2025 Feb 13:1-25.

[50] Adefabi A, Olisah S, Obunadike C, Oyetubo O, Taiwo E, Tella E. Predicting Accident Severity: An Analysis Of Factors Affecting Accident Severity Using Random Forest Model. arXiv preprint arXiv:2310.05840. 2023 Oct 9.

[51] Badidi E, Moumane K, El Ghazi F. Opportunities, applications, and challenges of edge-AI enabled video analytics in smart cities: a systematic review. IEEE Access. 2023 Aug 1;11:80543-72.

[52] Khushi M, Shaukat K, Alam TM, Hameed IA, Uddin S, Luo S, Yang X, Reyes MC. A comparative performance analysis of data resampling methods on imbalance medical data. IEEE Access. 2021 Aug 3;9:109960-75.

[53] Wang T, Lin Q. Hybrid predictive models: When an interpretable model collaborates with a black-box model. Journal of Machine Learning Research. 2021;22(137):1-38.

[54] 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. Int J Environ Res Public Health. 2022 Mar 2;19(5):2925. doi: 10.3390/ijerph19052925. PMID: 35270617; PMCID: PMC8910532.

[55] Rodríguez Hernández JM, Chaparro Narváez PE, Hidalgo Troya A, Piñeros Garzón FS. Undercounts stemming from misclassification derived from fatal injuries in traffic crashes in Colombia, 2010 to 2021. Traffic Injury Prevention. 2025 Apr 26:1-9.

[56] Aslam U, Fatima S. Dashcam Data Integrity in Digital Forensics: Exploring Video Sensors for Vehicular Investigations.

[57] Kang Y, Abraham J, Ceccato V, Duarte F, Gao S, Ljungqvist L, Zhang F, Näsman P, Ratti C. Assessing differences in safety perceptions using GeoAI and survey across neighbourhoods in Stockholm, Sweden. Landscape and Urban Planning. 2023 Aug 1;236:104768.

[58] Lindholm M, Richman R, Tsanakas A, Wüthrich MV. What is fair? Proxy discrimination vs. demographic disparities in insurance pricing. Scandinavian Actuarial Journal. 2024 Oct 20;2024(9):935-70.

[59] von Ungern-Sternberg A. Automated Law Enforcement-Perfect Vision or Dystopia?. Research Handbook on the Law of Artificial Intelligence, Edition (Elgar Forthcoming). 2024 Jul 19.

[60] Yigitcanlar T, Agdas D, Degirmenci K. Artificial Intelligence in local governments: perceptions of city managers on prospects, constraints and choices. Ai and Society. 2023 Jun;38(3):1135-50.