Accident Analysis and Prevention 197 (2024) 107457

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Predicting pedestrian-involved crash severity using inception-v3 deep learning model

Md Nasim Khan [a,*], Subasish Das [b], Jinli Liu [c]

[a] Senior Engineer, AtkinsRealis, 11801 Domain Blvd Suite 500, Austin, TX 78758, United States
[b] Assistant Professor, Texas State University, 601 University Drive, San Marcos, TX 78666, United States
[c] Geography and Environmental Studies, Texas State University, 601 University Drive, San Marcos, TX 78666, United States

ARTICLE INFO

Keywords:
Pedestrian
Crash Severity
DeepInsight
Inception-v3
Data Balancing
Synthetic Minority Oversampling Technique

ABSTRACT

This research leverages a novel deep learning model, Inception-v3, to predict pedestrian crash severity using data collected over five years (2016–2021) from Louisiana. The final dataset incorporates forty different variables related to pedestrian attributes, environmental conditions, and vehicular specifics. Crash severity was classified into three categories: fatal, injury, and no injury. The Boruta algorithm was applied to determine the importance of variables and investigate contributing factors to pedestrian crash severity, revealing several associated aspects, including pedestrian gender, pedestrian and driver impairment, posted speed limits, alcohol involvement, pedestrian age, visibility obstruction, roadway lighting conditions, and both pedestrian and driver conditions, including distraction and inattentiveness. To address data imbalance, the study employed Random Under Sampling (RUS) and the Synthetic Minority Oversampling Technique (SMOTE). The DeepInsight technique transformed numeric data into images. Subsequently, five crash severity prediction models were developed with Inception-v3, considering various scenarios, including original, under-sampled, over-sampled, a combination of under and over-sampled data, and the top twenty-five important variables. Results indicated that the model applying both over and under sampling outperforms models based on other data balancing techniques in terms of several performance metrics, including accuracy, sensitivity, precision, specificity, false negative ratio (FNR), false positive ratio (FPR), and F1-score. This model achieved prediction accuracies of 93.5%, 77.5%, and 85.9% for fatal, injury, and no injury categories, respectively. Additionally, comparative analysis based on several performance metrics and McNemar's tests demonstrated that the predictive performance of the Inception-v3 deep learning model is statistically superior compared to traditional machine learning and statistical models. The insights from this research can be effectively harnessed by safety professionals, emergency service providers, traffic management centers, and vehicle manufacturers to enhance their safety measures and applications.

1. Introduction

1.1. Background and problem statement

The escalating number of pedestrian-involved traffic crashes worldwide is a cause of growing concern, a situation that emphasizes the importance of immediate and effective measures to ensure pedestrian safety. In 2021, traffic crashes claimed the lives of 7,388 pedestrians, marking an upsurge of 12.5 percent in contrast to the previous year, 2020. Moreover, pedestrian injuries from traffic crashes were estimated at around 60,577 in the same year, a rise of 11 percent from 2020. Statistics from 2021 indicate that every 71 min a pedestrian lost their life, and every 9 min a pedestrian got injured in traffic crashes. Fatalities involving pedestrians represented 17 percent of all traffic-related deaths (NHTSA, 2021). Viewed on a global scale, pedestrian crashes hold the eighth position among the leading causes of death. Alarmingly, they are the foremost cause of mortality in young individuals aged between fifteen and twenty-nine. The World Health Organization (WHO) reported that each year, road traffic crashes claim approximately 1.3 million lives globally, leaving another 20 to 50 million individuals grappling with non-fatal injuries (WHO, 2023). Pedestrians are inherently vulnerable to traffic for a host of reasons. Unlike vehicle occupants, they lack physical barriers that provide protection during collisions, thereby escalating the risk of serious injuries or fatalities. Pedestrians,

* Corresponding author.
E-mail addresses: mdnasim.khan@atkinsrealis.com (M.N. Khan), subasish@txstate.edu (S. Das), jinli.liu@txstate.edu (J. Liu).

https://doi.org/10.1016/j.aap.2024.107457
Received 18 August 2023; Received in revised form 17 December 2023; Accepted 2 January 2024

especially children, the elderly, or those with disabilities, may not be able to move quickly enough to avoid an oncoming vehicle. In low-light conditions or at night, pedestrians often become virtually invisible to motorists. Moreover, distraction due to smartphones and impaired judgment from alcohol or drug use among pedestrians or drivers can increase the likelihood of a crash. Despite these daunting statistics and risk factors, it is crucial to note that enduring injuries or fatalities from pedestrian-involved motor vehicle crashes can be substantially reduced by ensuring prompt emergency medical intervention. Hence, the expeditious and accurate prediction of crash severity is especially important, paving the way for fast provision of emergency medical services and efficient traffic collision management.

The severity of pedestrian crashes has been extensively investigated in existing literature using traditional statistical approaches. A wide array of modeling techniques has been harnessed in earlier studies as discussed comprehensively in the literature review section. Despite their widespread use, these statistical methodologies carry predefined assumptions, and violation of these assumptions can lead to significant prediction inaccuracies when determining crash severity levels. In response to these limitations, machine learning has surfaced as a potent tool in recent years for predicting and analyzing the severity of pedestrian crash injuries, offering a means to address the inadequacies of conventional statistical methodologies. Although many studies have used machine learning, there exists a significant gap in research, with a clear lack of studies dedicated to the prediction of pedestrian-involved crashes that leveraged deep learning models. Such a model could potentially offer improved predictive performance compared to traditional statistical and machine learning models. Furthermore, while current state-of-the-art models can predict the severity of pedestrian crashes, there is still considerable room for improvement, particularly in ensuring balanced recalls across each severity level. This study, therefore, seeks to fill these gaps.

1.2. Research objective and significance

The primary objective of this study is to develop a robust model capable of predicting the severity of crashes involving pedestrians with superior and balanced accuracies across all crash severity categories. To achieve this goal, the study employs the advanced Inception-v3 deep learning model for predicting crash severity involving pedestrians. This study incorporates several interconnected phases of research. First, data relating to crash severity with pedestrian involvement is extracted and prepared from Louisiana. Subsequently, techniques like Random Under Sampling (RUS) and Synthetic Minority Oversampling Technique (SMOTE) are applied to balance the extracted crash severity data by adjusting the class distribution. Following this, a thorough exploration is conducted into the causes behind pedestrian-involved crashes, using the Boruta feature selection algorithm to select relevant variables. Subsequently, the research incorporates a robust method, DeepInsight, to transform numeric crash severity data into usable image data. This unique technique allows the representation of crash severity information in an image format, enabling the deep learning model to extract valuable patterns and features. Next, the study implements transfer learning with the Inception-v3 deep learning model, setting the stage for the construction of prediction models. Finally, the study compares the performance of the proposed model with other state-of-the-art statistical and machine learning methods.

While statistical models excel at providing interpretability and understanding the marginal effects of various risk factors associated with crash occurrences, they often fall short when it comes to developing highly accurate prediction models for real-world applications. In contrast, machine learning models can overcome the issues of low prediction accuracy generated by statistical models and can offer superior performance. However, the most cutting-edge advancement in the machine learning domain is the development of deep learning algorithms, which are primarily based on computer vision. However, pedestrian crash severity data is typically compiled and managed in a numeric format. To fully leverage the capabilities of advanced deep learning models, this study transformed the numeric data into image data to meet the input requirements of deep learning models.

Traditional machine learning models operate on feature vectors for prediction and assume that these features are independent, meaning their order does not significantly influence the prediction outcome. In contrast, deep learning models can harness non-linear relationships between features, consider the local spatial coherency of image pixels, and incorporate higher-order statistics of images. Deep learning models treat sequences of adjacent image pixels as dependent, thereby extracting valuable information during training that is not achievable with traditional models. While many advanced deep learning models are available, and further advancements in this domain are ongoing with the advancement of artificial intelligence and computer vision, this study employed the Inception-v3 deep learning model due to its ability to provide a high degree of accuracy while remaining computationally efficient.

It is important to note that the primary focus of this study was not to determine the superiority of the deep learning approach over statistical and traditional machine learning models. Instead, the aim was to illustrate how each method can offer valuable insights to researchers and transportation practitioners based on their intended objectives.

1.3. Research contribution

This study provides significant contributions and advancements over current methodologies. It represents one of the initial attempts to develop an advanced Inception-v3-based deep learning model for predicting pedestrian crash severity. Furthermore, the study employs the unique DeepInsight technique to convert numeric pedestrian crash severity data into image data. In contrast to previous studies that often concentrate on a limited set of factors, this research encompasses a broad range of variables, including crash, environmental, roadway, driver, pedestrian, and vehicle-related factors. The study also utilizes Boruta feature selection techniques to conduct a comprehensive examination of the factors contributing to crash severity involving pedestrians.

2. Literature review

A significant body of research has been devoted to understanding factors that lead to varying degrees of severity in pedestrian-vehicle crashes. Traditional statistical methodologies have often been the cornerstone of crash severity prediction in the research community. Olowosegun et al. (2022) focused on the elements that dictate the severity of pedestrian-motor vehicle accidents in Scotland between 2010 and 2018, with particular attention to junctions and crossings. Their approach leveraged correlated random parameter ordered probit models to factor in unobserved variations. The findings pointed to several influential elements such as road, location, weather, vehicle, driver traits, and timing factors. Differences were noted between signalized and unsignalized junctions and crossings controlled either physically or by humans, particularly in weather, hazards, and time factors. Zamani et al. (2021) utilized the Los Angeles crash data from 2012 to 2017 to identify the determinants of pedestrian injury severity. Using a random parameters logit model, they explored key variables and their constancy over time. They concluded that the influential variables were not static over the seven-year study period, underscoring the necessity for dynamic analysis of crash data. Li et al. (2021) explored pedestrian-vehicle crash data in North Carolina between 2007 and 2018 to assess varying factors influencing pedestrian injury severity. Utilizing random parameters logit models, they found that factors like ambulance rescue and curved roadway consistently impacted pedestrian injury severity while others exhibited strong temporal variations. Kitali et al. (2018) explored risk factors related to injury severity in pedestrian-vehicle accidents involving older pedestrians. Their model, utilizing a Dirichlet random-effect logistic model (DRL), managed to account for unobserved variations in crash data and outperformed conventional models, achieving 90 % accuracy. Ukkusuri et al. (2011) conducted a study on pedestrian crash frequencies in New York City. Their model, a random parameter negative binomial model, managed to predict crash frequencies by incorporating unobserved variations across spatial zones. Guo et al. (2020) introduced a two-level random intercept model using Bayesian probability inference for predicting pedestrian crash severity. The model's utility was demonstrated using pedestrian crash data from Colorado, and results suggested improved prediction accuracy compared to traditional methods. Wang et al. (2020) focused on factors relating to pedestrian red-light violations and injury severity in pedestrian-motor vehicle accidents at signalized crossings using random parameter probit models. Their model managed to account for individual-specific variations, and the findings indicated a need for countermeasures addressing these safety concerns. Bhat et al. (2017) developed an innovative spatial random coefficients flexible multivariate count model to explore pedestrian injuries, accounting for different risk factors for various injury types, and addressing excess zeros and spatial dependencies. Kwayu et al. (2020) aimed to devise a methodology for creating pedestrian safety performance functions (SPFs) to evaluate pedestrian traffic crash risks. They conducted a study on all urban intersections in Michigan State, utilizing a Zero-Inflated Poisson (ZIP) model for formulating the final pedestrian SPFs. Song et al. (2020) explored pedestrian-vehicle crashes in North Carolina between 2007 and 2018 using spatiotemporal analysis and hierarchical Bayesian random-effects models. The study identified significant temporal and spatiotemporal instability of factors affecting pedestrian injuries. A recent study utilized spatiotemporal logistic models to delve into the severity of injuries from motor vehicle–pedestrian collisions at urban junctions (Zeng et al., 2023). The most effective model was found to be the one employing Leroux conditional autoregressive prior and random walk structure, as indicated by goodness-of-fit and classification accuracy metrics. Parameters like pedestrian age, head injury, pedestrian location and activity, driver maneuvers, vehicle type, initial point of collision, and traffic congestion status considerably influenced the severity of pedestrian injuries. Another study by Adanu et al., (2023) implemented a random parameter multinomial logit model to pinpoint crash variables that significantly correlated with child pedestrian crash outcomes. The findings suggested that the chances of children getting fatally injured in crashes increased when drivers were speeding and not attentive.

Over the past few years, machine learning has evolved as a promising method for crash injury severity predictions, providing an alternative to traditional statistical approaches that may be limited in certain aspects. A study by Zhang et al. (2018) assessed the efficacy of four machine learning techniques, namely, K-Nearest Neighbor (KNN), Decision Tree (DT), Random Forest (RF), and Support Vector Machine (SVM). Another study by Govinda et al. (2022) utilized Support Vector Machines (SVM) to evaluate the severity level of interactions between pedestrians and vehicles at intersections without controls. The findings from SVM indicated that, with consistent vehicle and pedestrian characteristics, the critical Interaction Index (RI) value for serious incidents tended to drop as the speed of pedestrian crossings accelerated. The study concluded that machine learning models, despite a potential for overfitting, generally offer superior prediction accuracy compared to statistical models. Wu et al. (2023) emphasized the necessity for context-rich data and severity levels in automated transport systems, especially in mixed traffic settings involving vehicle–pedestrian conflicts. The research employed the CRSS dataset, which contains police-reported crashes, and used the XGBoost algorithm for feature identification. The XGBoost method was found to surpass random forest in managing crash data and provided consistent features in alignment with prior research. Factors such as speed limit, light conditions, pre-crash movements, pedestrian location, driver distraction, and impairment were all revealed as critical contributors to crash severity. Goswamy et al. (2023) compared the efficacy of Rectangular Rapid Flashing Beacons (RRFB) on crash severity using XGBoost and Random Parameters Discrete Outcome Models (RPDOM), analyzing data from 312 pedestrian crossing locations. The study demonstrated that RRFB positively impacted the reduction of nighttime crashes, with XGBoost outperforming RPDOM in predicting nighttime crash severity. Toran Pour et al., (2017) established three models using distinct decision trees to identify factors influencing pedestrian crash severity. They applied bagging and boosting techniques to enhance the accuracy, stability, and robustness of decision trees. The results indicated that boosting improved the accuracy of individual decision tree models by 46 % and underscored the significance of neighborhood social characteristics alongside traffic and infrastructure variables in determining pedestrian crash severity. Zhu et al. (2023) utilized latent class clustering analysis in conjunction with RF to investigate into the factors associated with pedestrian and bicyclist involved crashes. Their results indicated that crashes involving lower-income drivers and non-white victims were more likely to occur in areas with higher pedestrian/bicyclist exposure, faster speed limits, and wider roads. Kong et al. (2023) adopted an interpretable machine learning framework using SHapley Additive exPlanations (SHAP) to understand the factors associated with critical pedestrian-involved near-crash events. The results suggested that pedestrians with a higher walking speed and areas with high pedestrian volume were more prone to critical near-crash incidents. Zhu et al. (2022) proposed a Vehicle-Pedestrian Detection (VPD) algorithm based on a Convolutional Neural Network (CNN) to enhance prediction accuracy and the recognition of interactions between autonomous vehicles and pedestrians. Their algorithm improved the prediction accuracy by at least 1.94 % compared to advanced CNNs. A study by Wen et al. (2021), identified imbalanced data as a major challenge in crash severity modeling. They recommend employing advanced machine learning techniques such as Graph Convolutional Network (GCN) and exploring interpretable machine learning to interpret results and identify causal relationships. Furthermore, a study by Abdelwahab and Abdel-Aty (2001) found Artificial Neural Networks (ANN) to perform better than ordered logit models in crash severity prediction. Meanwhile, a study by Alkheder et al. (2017) used the k-means clustering technique to categorize crash severity data into subsets and applied ANN to model them separately, which significantly improved prediction performance. Another study by Mokhtarimousavi et al. (2020) used ANN to investigate the nonlinear relationship between explanatory variables and severity outcomes in vehicle–pedestrian crashes, finding that optimized ANN outperformed traditional statistical methods.

Numerous safety studies have conducted comparisons between machine learning models and traditional statistical models in terms of their predictive performance. For instance, (Zhang et al., 2018) examined two commonly used statistical methods, the ordered probit model and multinomial logit model, alongside four popular machine learning methods, namely KNN, DT, RF, and SVM. Their findings indicated that machine learning methods generally exhibited higher predictive accuracy compared to statistical methods, although they were susceptible to overfitting issues. The RF method displayed the best overall prediction performance, particularly in severe crash scenarios, whereas the ordered probit method yielded the weakest results. In another investigation, (Iranitalab and Khattak, 2017) compared the effectiveness of four statistical and machine learning methods, including Multinomial Logit (MNL), Nearest Neighbor Classification (NNC), SVM, and RF, in predicting traffic crash severity. Their results indicated that NNC exhibited the best predictive performance for overall and more severe crashes, with high prediction accuracies. Another study by (Jamal et al., 2021) argued that traditional statistical models, often characterized by predefined correlations and intrinsic assumptions, could lead to biased predictions when these assumptions were violated. Their analysis demonstrated significantly superior prediction performance of XGBoost when compared to logistic regression. (Hosseinzadeh et al., 2021) conducted a comparative analysis involving SVM and the random parameter binary logit model (RPBL) to investigate the factors influencing the severity of large truck-involved crashes. Their findings emphasized the importance of a complementary approach, utilizing both parametric RPBL and non-parametric SVM, to identify the primary contributing factors affecting crash severity in these situations. In a study by (Wang and Kim, 2019), a comparison was made between the multinomial logit (MNL) model and the RF model for predicting and identifying factors influencing crash severity. The research concluded that RF demonstrated higher predictive accuracy than MNL, even though the differences were not significant. Sensitivity analysis further indicated that RF was less sensitive compared to MNL. Moreover, a study by (Morris and Yang, 2021) explored the predictive performance of three ensemble machine learning models, specifically CatBoost, XGBoost, and RF, in comparison with the classic statistical model Nested Logit. Their research reported that tree ensemble methods, including CatBoost and XGBoost, demonstrated superior performance compared to the classic Nested Logit model while remaining highly interpretable with the aid of SHAP.

Based on the review of the literature, it becomes evident that machine learning models often outperform statistical models in terms of prediction performance. However, this does not necessarily imply that machine learning models are superior in every aspect. Statistical models possess their own set of advantages, most notably their superior interpretability. Statistical models excel at providing a clear understanding of the marginal effects associated with various risk factors influencing crash occurrences. Therefore, the choice of model should be contingent upon the specific objectives of the study. If the primary goal is to achieve interpretability in the context of a safety study, then a statistical model is recommended. On the other hand, if the primary objective is to enhance crash prediction performance, then machine learning models are the preferred choice.

3. Data preparation

This study employs crash data that encompasses a period of five years, specifically from 2016 to 2021, and incorporates records of pedestrian-involved accidents in the state of Louisiana. In the study, the initial data set contained fifty-six variables. However, not all variables were relevant to the analysis of pedestrian-involved crashes. The researchers focused specifically on variables that were pertinent to pedestrian incidents. To narrow down the variables from the initial fifty-six, correlation analysis, feature importance analysis, or domain expertise were employed to determine the variables that had the strongest relationships with pedestrian incidents.

To minimize noise within the data and simplify interpretation, the categories within each variable were refined, limiting them to a maximum of six categories. In cases where a variable contained more than six categories, only the top five most frequent were retained, with the remainder being consolidated under the category ’other’. In the context of this study, the decision to limit the categories within each variable to a maximum of six was made after conducting several iterations. This process has several advantages. First and foremost, this approach enhances the simplicity and interpretability of the data. By reducing the number of categories, the relationships between variables become more straightforward and easier to analyze, allowing for clearer interpretations of the findings. Moreover, limiting the categories helps prevent overfitting, a common issue in statistical modeling, where the model captures noise instead of meaningful patterns. Each crash record in the dataset represents an individual-level entry, thereby accounting for the unique characteristics and outcomes of each person involved in a particular incident. It is important to acknowledge that crash scenarios can encompass multiple vehicles and pedestrians, each of whom may experience varying degrees of severity. Given this variability, individual-level crash data was employed in this study to accurately capture the severity experienced by each person involved. The dataset incorporates 8213 instances of pedestrian-involved crashes.

The pedestrian crash severity data from 2017 to 2021 reveals significant insights into the safety landscape (see Table 1). On average, there were approximately 147 fatal crashes annually, with a notable increase observed in 2021. Incapacitating injury crashes averaged around 175 per year, contributing to a total of 874 over the five-year span. The non-incapacitating injury category demonstrated a fluctuating trend, with an average of about 594 crashes each year. Minor injury crashes, averaging approximately 556 annually, resulted in a total of 2,780 incidents over the five years. Conversely, crashes with no reported injuries occurred at an average rate of 171 per year, totaling 855 over the entire period.

The final dataset included forty variables associated with pedestrian, location, weather, driver, road, and vehicle characteristics. Pedestrian crash severity was compartmentalized into three categories: fatal, injury, and no injury. Injury crashes were notably more prevalent than the other two categories, with 6624 recorded incidents, as opposed to 734 fatal crashes and 855 incidents with no injuries. The categorization of crash severity is a multifaceted process, often involving the subdivision of severity levels into various clusters. For instance, some studies group fatal and incapacitating injuries together, non-incapacitating and minor injuries as a separate cluster, and no injury in another category. Alternatively, other research adopts a classification scheme where fatal, incapacitating, and non-incapacitating injuries form one group, minor injuries are grouped together, and no injury is treated as a distinct category. The current study divided pedestrian crashes into two main groups: fatal and non-fatal. For the non-fatal ones, two additional groups were considered: those with injuries and those without any injuries. This approach is able to determine the critical distinction between outcomes with severe consequences (fatal incidents) and those resulting in varying degrees of injury or no injury. It is important to know that a more detailed classification into five distinct levels might result in sparse data in certain categories, potentially affecting the prediction accuracy of the modeling technique. By consolidating injury levels into a single ‘Injury’ category, this study aimed to strike a balance between meaningful distinctions and analytical feasibility.

Table 2 illustrates the descriptive statistics of the key variables used in the models. Pedestrian impairment was segmented into five categories, with roughly 5.2 % of pedestrians involved found to be impaired by alcohol. Regarding driver impairment, approximately 3.4 % were impaired by alcohol. To assess pedestrian attentiveness and distractions, this study relied on structured data provided by the Louisiana Department of Transportation and Development (LaDOTD). Within this dataset, a variable called ‘Pedestrian Condition’ was utilized, which included specific categories such as ‘inattentive’ and ‘distraction.’ Around 22 % of the pedestrians are either distracted or inattentive. Inattentiveness refers to a situation where the pedestrian is not focusing on the task of walking or crossing the road. This type of distraction is considered cognitive, making it challenging for an officer to pinpoint the exact reason for the pedestrian’s inattentiveness, especially if their attention is occupied by a mental task unrelated to walking or crossing the road. While articulating why a pedestrian was inattentive might be difficult, there could still be apparent evidence suggesting the pedestrian’s inattentiveness. Similarly, a distracted pedestrian is one who is actively involved in an activity diverting their attention from the task of walking or crossing the

Table 1
Yearly pedestrian crash counts by crash severity.

Crash Severity Level	2017	2018	2019	2020	2021	Total
Fatal (K)	117	164	122	146	185	734
Incapacitating Injury (A)	160	175	179	176	184	874
Non-incapacitating Injury (B)	678	668	617	495	512	2970
Minor Injury (C)	652	624	602	453	449	2780
No Injury (O)	166	178	200	143	168	855
Total	1773	1809	1720	1413	1498	8213

Table 2
Descriptive statistics of the key variables.

Variable	Categories	Count	%	Variable	Categories	Count	%	Variable	Categories	Count	%
Crash Severity (Severity)	Fatal	734	8.94	Pedestrian Impairment (Ped_Impair)	No	5146	62.66	Pedestrian Action (Ped_Act)	Crossing_Int	1882	22.91
	Injury	6624	80.65		Unknown	2492	30.34		Crossing_seg	1778	21.65
	No Injury	855	10.41		Yes_Alcohol	427	5.2		NotRoadway	653	7.95
Pedestrian Age (Ped_Age)	Children/ Infant	725	8.83		Yes_Both	88	1.07		Other	1828	22.26
	Middle	4919	59.89		Yes_Drug	60	0.73		Unknown	775	9.44
	Old	559	6.81	Pedestrian Condition (Ped_Cond)	AlcImpaired	381	4.64		Walking	1297	15.79
	Unknown	396	4.82		Distracted	181	2.2	Crash Type (CrashType)	LeftTurn	228	2.78
	Young	1614	19.65		Inattentive	1612	19.63		NCWMV	5597	68.15
Pedestrian Residency (Ped_State)	LA	7378	89.83		Normal	3324	40.47		Other	1210	14.73
	Non-LA	835	10.17		Other	399	4.86		RearEnd	343	4.18
Pedestrian Gender (Ped_Gen)	Female	2816	34.29		Unknown	2316	28.2		RightAngle	528	6.43
	Male	5160	62.83	Pedestrian Race (Ped_Gen)	Black	4364	53.14		Sideswipe	307	3.74
	Unknown	237	2.89		Indian	4	0.05	Crash Location (RoadRel)	Median	18	0.22
Day of Week (DOW)	Weekday	6022	73.32		Other	261	3.18		OnRoadway	7339	89.36
	Weekend	2191	26.68		Unknown	272	3.31		Other	213	2.59
Season of Year (Season)	Autumn	2280	27.76		White	3312	40.33		Shoulder	643	7.83
	Spring	2034	24.77	Vehicle Movement Reason Prior to Crash (MovReason)	DriverCondition	316	3.85	Vehicle Movement Prior to Crash (PriorMov)	Lanechange	74	0.9
	Summer	1700	20.7		DriverViolation	1358	16.53		LeftTurn	725	8.83
	Winter	2199	26.77		NormalMovement	3771	45.92		Other	1628	19.82
Most Harmful Event (MostHarm)	MVinTrans	1872	22.79		Other	604	7.35		Stopped	186	2.26
	Other	5994	72.98		Unknown	1753	21.34		StraightAhead	4864	59.22
	ParkedMV	144	1.75	Hit and Run (HAR)	No	5906	71.91		Unknown	736	8.96
	Unknown	203	2.47		Yes	2307	28.09	Land Use (Loc)	Business	4900	59.66
Involvement of truck/Bus (TrkBus)	No	7928	96.53	Access Control (Access)	Full Control	344	4.19		Industrial	148	1.8
	Yes	285	3.47		No Control	7052	85.86		Other	542	6.6
Alcohol Involvement (ESTAlc)	No	6939	84.49		Other	105	1.28		Residential	2515	30.62
	Yes	1274	15.49		Partial Control	712	8.67		School	108	1.31
Run of Road (RwD)	No	798	9.72	Alignment (Align)	Curve	243	2.96	Road Type (RoadType)	OneWay	1025	12.48
	Yes	7415	90.28		Other	214	2.61		Other	155	1.89
Lane Departure (LaneDepart)	No	760	9.25		Straight	7756	94.44		TwoWayDiv	1834	22.33
	Yes	7453	90.75						TwoWayUndiv	4944	60.2
Highway Type (HwyType)	CityStreet	4102	49.95	Posted Speed Limit (PSL)	25MPHorLower	3129	38.1		TwoWayWithBarrier	255	3.1
	Other	387	4.71		30-45MPH	3938	47.95	Traffic Control Device (TCD)	NoControl	1950	23.74
	ParishRoad	1140	13.88		50-60MPH	949	11.55		Other	2194	26.71
	StateHwy	1682	20.48		65MPHandAbove	197	2.4		StopSign	737	8.97
	U.S.Hwy	902	10.98	Number of Occupants (NumOcc)	Multiple	860	10.47		WhiteDashedLine	1797	21.88
Weather Condition (Weather)	Clear	6467	78.74		Single	6268	76.32		YellowDashedLine	745	9.07
	Cloudy	1064	12.96		Two	1085	13.21		YellowNoPassingLine	790	9.62
	Fog/Smoke	66	0.8		Multiple	860	10.47	Driver Condition (DrCond)	Distracted	224	2.73
	Other	84	1.02	Visibility Obstruction (VisObs)	MovingVehicles	94	1.14		Inattentive	1237	15.06
	Rain	524	6.38		NoObscurements	5337	64.98		Normal	4084	49.73
Driver Impairment (Driver_Impair)	No	5439	66.22		Other	482	5.87		Other	512	6.23
	Unknown	2374	28.91		Rain/ SnowWinshield	155	1.89		Unknown	2156	26.25
	Yes_Alcohol	281	3.42		Unknown	2145	26.12	Driver Distraction (DrDistract)	No	4492	54.69
	Yes_Both	84	1.02	Driver Gender (DrGen)	Female	2632	32.05		Unknown	3282	39.96
	Yes_Drug	35	0.43		Male	3891	47.38		Yes_Inside	181	2.2
Vehicle Headlight (VehLight)	DRL	508	6.19		Unknown	1690	20.58		Yes_Outside	258	3.14
	Off	1364	16.61	Vehicle Condition (VehCond)	Defective	76	0.93	Traffic Rule Violation (Violation)	DisregardedTrafCont	128	1.56
	On	3398	41.37		Good	1	0.01		FailuretoYield	579	7.05
	Unknown	2943	35.83		NoDefect	5880	71.59		ImproperBacking	96	1.17
Driver Race (DrRace)	Black	3152	38.38		Other	136	1.66		NoViolations	3572	43.49
	White	3119	37.98		Unknown	2120	25.81		Other	2558	31.15
	Other	1942	23.65	Vehicle Type	Car	3516	42.81		Unknown	1280	15.59
Lighting Condition (Lighting)	DarkNoStLt	1222	14.88		LightTruck	2023	24.63	Driver Age (DrAge)	Middle	4390	53.45
	DarkStLt	2730	33.24		Med/LargeTruck	201	2.45		Old	778	9.47
	Daylight	3940	47.97		Motorcycle	56	0.68		Unknown	1807	22
	Dusk/Dawn	239	2.91		Other	588	7.16		Middle	4390	53.45
	Other	82	1		SUV	1829	22.27				

road. Distractions can be manual, visual, or cognitive in nature, indicating activities that involve the hands, eyes, or mind, respectively. Pedestrian ages were sorted into five categories: children/infants, middle-aged, old, young, and unknown. Approximately 19.6 % of pedestrians were classified as young, while 8.8 % were either children or infants. Land use was sorted into five categories: business, industrial, other, residential, school, and other. Approximately 59.7 % of crashes occurred in business districts, and 30.6 % occurred in residential areas. Concerning posted speed limits, approximately 61.9 % of pedestrian-involved crashes happened on roadways with speed limits above 30 mph. Lighting conditions were categorized into five groups: dark without streetlights, dark with streetlights, daylight, dusk/dawn, and other. Approximately, 48.1 % of pedestrian-involved crashes occurred during nighttime.

4. Methodology

4.1. Data balancing

The dataset employed for this study demonstrated a significant imbalance as over 80 % of the data related to pedestrian-involved incidents fell within the injury crash severity category. In the realm of machine learning algorithms, an imbalanced dataset refers to a situation where the classification classes exhibit a notable uneven distribution (Chawla, 2009). The classes with less representation, in contrast to their counterparts, are characterized as “Minority Classes (MIC)”. The extent of data imbalance can be stratified into three categories, namely, mild (20 %-40 % MIC), moderate (1 %-20 % MIC), and extreme (less than 1 % MIC) (Khan and Ahmed, 2023; Google, 2021). As highlighted in the previous discussion, a comparison of 6624 recorded incidents with 734 fatal and 855 non-injury crashes revealed a considerable imbalance within the dataset. Both the fatal and non-injury categories were notably underrepresented, falling into the MIC. It is critical to note that this scarcity of data from the MIC could potentially lead to inaccurate interpretations and undermine the reliability of the crash severity prediction.

While the overall crash severity prediction might be high due to the accurate classification of the predominant sample (in this case, injury crashes), the MIC might not be precisely predicted. Consequently, the dataset requires balancing before initiating the analysis, which can be executed via under sampling, oversampling, or a combination of both techniques. The under-sampling approach eliminates multiple samples from the “Majority Classes (MAC)” to align with the MIC, while the oversampling method creates artificial MIC data points to attain balance. The process of under sampling was performed using RUS, which involves random selection of data from the MAC to correspond with the MIC. Post under sampling, the dataset comprised 2022 pedestrian-involved crash instances, with 734 instances for each crash severity category. On the other hand, oversampling was implemented via a method called SMOTE, a commonly used oversampling technique to address class imbalance in machine learning. It involves creating synthetic points that lie between existing minority class samples. The SMOTE algorithm begins by selecting a random minority class sample from the dataset and identifying its k nearest neighbors in the feature space. It then fabricates synthetic points through interpolation between the minority sample and its k nearest neighbors. This is realized by selecting a random neighbor from the k nearest neighbors, computing the difference between the feature values of the neighbor and the minority sample, multiplying the difference by a random number between 0 and 1, and adding it to the minority sample. The end product is a new synthetic point that lies along the line connecting the minority sample and its neighbor in the feature space. This synthetic point generation process is repeated until the desired level of oversampling is reached, and the new synthetic points are added to the minority class, thereby augmenting its size within the dataset (Chawla et al., 2002). Post oversampling, the total data resulted in 19,872 pedestrian-involved crash instances, with 6,624 instances for each category. Lastly, a combination of RUS and SMOTE was applied to balance the data, yielding 12,000 crash instances with 4,000 instances per category. Note that the under sampling approach may induce biases as critical information could be discarded in the process. Conversely, the oversampling technique could trigger overfitting issues (Ganganwar, 2012). Therefore, both under sampling and oversampling techniques were utilized and compared in this study based on their respective advantages and disadvantages.

4.2. Variable importance

In this study, the Boruta feature selection algorithm is utilized to uncover and scrutinize the key variables associated with pedestrian-involved crashes. Boruta, an algorithm rooted in Random Forest (RF) classification, adopts an innovative procedure involving duplicates of original variables, designated as Shadow Features (SF). Each data point in the dataset is arbitrarily assigned an SF, leading to the formulation of decision trees reliant on these SFs. A distinctive feature of the Boruta method is the way it determines the relative importance of a variable. It quantifies the reduction in the model’s accuracy attributable to the SF, which subsequently becomes the basis for calculating the Z scores. These scores, which are computed from the mean and standard deviation of the accuracy loss, serve as a measure of significance within the Boruta framework. The significance set of the SFs acts as an essential benchmark to discern the relevance or importance of the original variables in the crash dataset. Subsequently, the relevance of the original variables is evaluated against the highest significance of the SFs. This comprehensive process lies at the heart of the Boruta feature selection algorithm and consists of several pivotal stages.

The first stage involves expanding the information system by crafting SF for each original variable, ensuring a minimum of five SF per variable. Following this, RF models are run on each of the SF, after which the associated Z scores are determined. Once this is accomplished, the maximum Z score among the shadow attributes (MZSF) is derived. The MZSF acts as a comparative standard against which the significance of each original variable is gauged. Depending on this comparison, the variables are split into two categories. Variables exhibiting significantly lower importance than the MZSF are discarded as Unimportant, while those with markedly higher importance than the MZSF are earmarked as Important. In the final stage, all the SF are eliminated, and the entire process is reiterated. This cycle is maintained until all original variables have been assigned their respective importance levels. Consequently, through its systematic approach, the Boruta algorithm provides a powerful means to identify and evaluate the relevance of variables in a dataset, specifically those related to pedestrian-involved crashes (Ahmadpour et al., 2021; Das et al., 2023).

4.3. Conversion from numeric to image data

Pedestrian crash severity data are typically compiled and managed in a numeric format. This study utilized a robust technique known as DeepInsight to convert numeric attributes from a crash database into an image-based format (Sharma et al., 2019). This transformation was required to accommodate the pre-trained CNN model Inception-v3, which only accepts images as input. Traditional machine learning models operate on feature vectors for prediction and assume these features to be independent, meaning their order does not significantly influence the prediction outcome. However, CNNs are capable of harnessing the non-linear relationships of features, acknowledging the local spatial coherency of image pixels, and employing higher-order statistics of images. CNNs treat the sequence of adjacent image pixels as dependent, a capability not shared with conventional machine learning models, thus extracting useful information during training that is not achievable with traditional models.

DeepInsight’s underlying principle involves converting a numeric feature vector (pedestrian crash data variables, in this case) into a feature matrix that retains the feature positions. Once this matrix is verified, it can be filled with feature values, generating a single image for every crash instance. Fig. 1 depicts the key steps involved in transforming numeric crash data into images. The final dataset is structured such that the rows represent pedestrian-involved crash instances and columns indicate features or variables. The transformation process begins with transposing the crash severity data to fit its input specifications, with each row signifying a feature and each column a crash instance. Subsequently, a dimensionality reduction method known as t-Distributed Stochastic Neighbor Embedding (t-SNE) was applied to the transposed data to map crash severity data onto a 2D plane. Here, each point in the plane denotes a feature’s Cartesian coordinates. These points only represent the locations, not the values, of the features. Notably, t-SNE is predominantly used to visualize high-dimensional

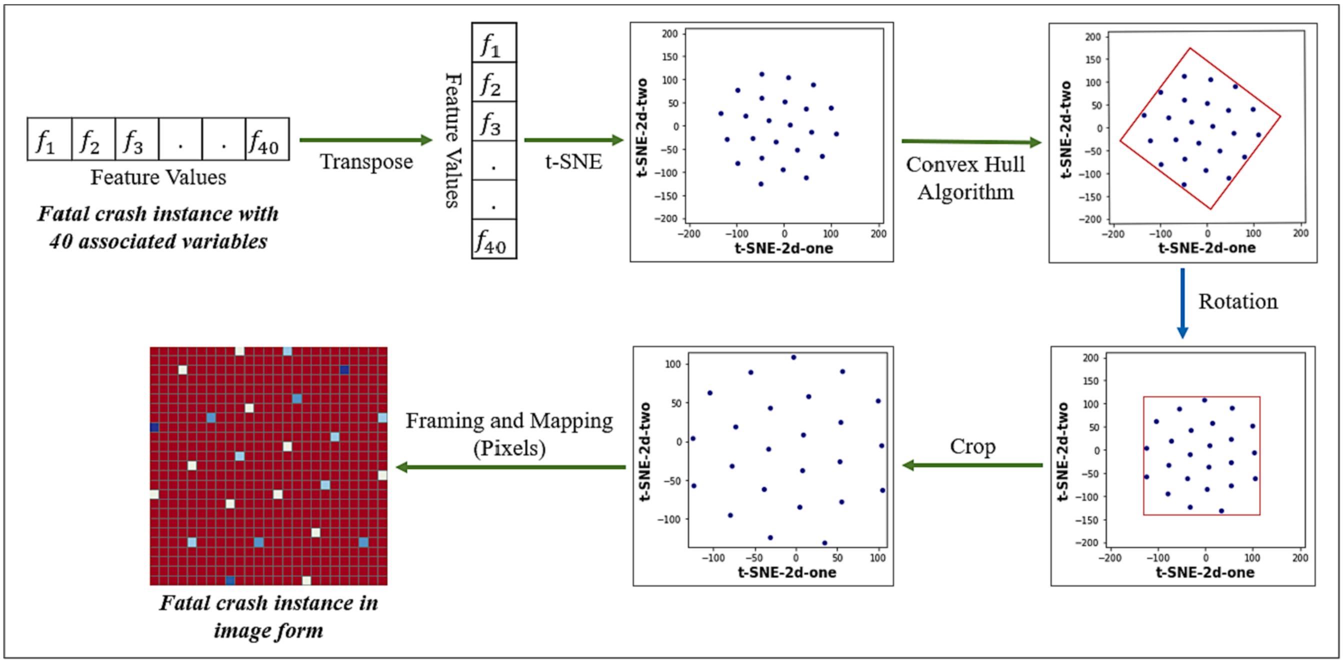

Fig. 1. Illustration of numeric to image transformation with a fatal crash instance.

datasets. It works by reducing the divergence between two distributions: one measuring the pairwise similarities of the input data and the other measuring the pairwise similarities of the corresponding points in the lower-dimensional embedding (van der Maaten and Hinton, 2008).

After mapping the features, the convex hull algorithm was implemented to determine the smallest rectangle encapsulating all points (features), as illustrated in Fig. 1. The points and the rectangle were then rotated for a horizontal alignment and unnecessary spaces were cropped out. Subsequently, the Cartesian coordinates were converted into pixel format by normalizing the coordinates and multiplying them with a predefined image height (Khan et al., 2024; Rahim and Hassan, 2021). The pixel frame accounted for feature locations for a sample x_i where $i = 1, 2, 3, \cdots, n$. Upon establishing feature locations, feature values were assigned to the corresponding pixel locations. It is crucial to mention that if multiple features overlapped, their corresponding features were averaged and placed in the same pixel location (Sharma et al., 2019). The image resolution influences the number of overlapping features. Therefore, to ensure an accurate image representation, an image size compatible with the feature count should be selected. Given the number of features in the crash severity database, an image size of 50×50 pixels was deemed suitable.

4.4. Deep learning Algorithm: Inception-v3

This study considered a robust and state-of-the-art deep pre-trained Convolutional Neural Network (CNN) called Inception-v3 to train and validate the pedestrian crash severity prediction models. This architecture has been selected based on the ability to obtain a high amount of prediction accuracy. The Inception-v3 model is a type of CNN architecture, renowned for its diverse applications in computer vision tasks (Dong and Zhang, 2019). First unveiled by Szegedy et al. in 2016 (Szegedy et al., 2016), this model is essentially an evolution of the conventional CNN, which was primarily designed to tackle image recognition problems. A standard CNN, mimicking the human process of image recognition, usually encompasses five types of layers: input, convolutional, pooling, fully connected, and output. To address more complex problems, this structure can be expanded to include additional layers. Before the inception of Inception-V3, Google had introduced GoogleNet in 2014 - a deeper CNN that employed the Inception network structure (Sam et al., 2019). GoogleNet was celebrated for its depth of analysis and reduced network parameters, making it a potent model for image classification tasks. The Inception-V3 model, building upon its predecessors, incorporates several notable design concepts. Among these are the use of inception modules, factorization, and auxiliary classifiers (Szegedy et al., 2015). An inception module, which serves as the cornerstone of the network, carries out multiple convolutions of varying sizes in parallel, followed by the concatenation of their outputs. This approach enables the network to seize both local and global information. Factorization helps lessen the computational cost by replacing larger convolutions with combinations of smaller ones.

The Inception architecture is characterized by its ability to scale down its dimensions based on the computational effort required to process the network. Generally, an Inception model has three distinct sizes of convolution layers and a single max pooling layer. After each layer performs convolution, the results are aggregated and then combined to form the deep learning structure. This structure has undergone several modifications over time, resulting in different versions of the model. Inception-v3, for instance, brought significant enhancements, such as the implementation of batch normalization, further factorization, and the adoption of the Root Mean Square Propagation (RMSProp) optimizer.

The training process of the Inception-v3 model typically involves transfer learning, a technique in which a model that was initially pre-trained on a comprehensive dataset is fine-tuned to address a specific task. By leveraging the representations learned during the pre-training phase, transfer learning can enhance the model's performance on the target task, minimizing the need for extensive training data. Existing studies suggest that to classify a new image using the Inception-v3 model, one can modify the architecture of the fully connected layers while retaining the parameters of all convolution layers (Meena et al., 2023). In the context of this study on pedestrian crash severity prediction, the Inception-v3 model was adapted using the same transfer learning approach. Instead of its original use-case for general image categorization, it was adjusted to process images representing instances of pedestrian-involved crashes. Modifications to the architecture involved replacing the original input layers and the final three layers with a new input layer and a fully connected layer. These changes were implemented to accommodate the specific input (i.e., transformed images of crash instances), and to align with the output specifications of this study (i.e., categorizing crash severity into fatal, injury, and no injury classes). Additionally, a softmax layer and an output layer were incorporated into the model. The softmax layer, specifically, is key for multi-class classification problems like ours, as it provides a set of probabilities for the potential output categories. Given our three-category output, the softmax layer ensures that the sum of the probabilities for 'fatal', 'injury', and 'no injury' is equal to one for each crash scene image. It is crucial to note that to meet the input requirements of the Inception-v3 model, the images were resized to dimensions of 299 × 299 × 3. As illustrated in Fig. 2, the overall architecture of the Inception-v3 model features a sequential combination of convolution blocks, inception modules, and classifiers, culminating in the final output. This configuration maintains the conventional structure of CNNs designed for image categorization, thus ensuring an efficient and accurate approach to pedestrian crash severity prediction.

5. Results and discussion

5.1. Validation and parameter tuning

The process of optimizing the pedestrian crash severity prediction model involved fine-tuning the parameters and training options of the Inception-v3 models. The dataset was initially partitioned into two segments: 80 % was designated for training and validation purposes, while the remaining 20 % was reserved for testing the performance of the trained models. It is important to emphasize that this 20 % subset was exclusively for testing and was never incorporated into the training or validation stages. Subsequently, 80 % of the data allocated for training and validation was further divided, with 80 % serving for training and 20 % designated for holdout validation. In the training phase, the models were validated every ten iterations. A critical part of the optimization was minimizing the loss function, also referred to as the cost, which signifies the predictive capability of the model. Lower costs denote better-performing models (Khan and Ahmed, 2020). For this study, three distinct optimizers were employed: Stochastic Gradient Descent with Momentum (SGDM), Root Mean Square Propagation (RMSProp), and Adaptive Moment Estimation (Adam), with SGDM providing the most accurate predictions across all models. Other parameters like the initial learning rate, learning rate drop factor, maximum epochs, batch size, and L2 Regularization factor were also adjusted. It's important to remember that parameter tuning doesn't have a universal solution but is mainly experiment-based. In accordance with previous research, this study utilized a method known as grid-search for fine-tuning hyperparameters and training options (Chicco, 2017).

Table 3 provides a summary of the training and validation details for the pedestrian-involved crash severity prediction models that were developed. It's important to note that during the data preparation stage, the numeric crash records were transformed into images to meet the input requirements of the Inception-v3 deep learning models. A detailed methodology for this conversion is described in the methodology section. Subsequently, after the conversion, the data in image format was used to train the models. However, the number of images for each model varied depending on the data balancing technique employed.

To optimize computational resources, the grid search was limited to the parameters with a substantial impact on model performance. Through a series of experiments, the ideal parameter values and training options were identified as follows: the initial learning rate was reduced to 0.001 from the original 0.01, the learning rate drop factor was elevated to 0.5 from 0.1, the learning rate drop period was adjusted to 8 from 10, the factor for L2 Regularization was increased to 0.008 from 0.0001, the maximum epochs were cut down to 15 from 30, and the

Table 3
Validation Summary of the Trained Pedestrian Crash Severity Prediction Models.

Model Name	Data Used	Number of Images	Number of Iterations	Validation Accuracy (%)	Training Time (Min)
MNB	Original imbalanced data	8,213	780	84.4	19.9
MBU	Balanced data using under sampling	2,202	210	66.4	4.6
MBO	Balanced data using over sampling	19,872	1905	87.0	63.2
MBB	Balanced data using both under and over sampling	12,000	1140	86.1	31.9
MBR	Balanced data using both under and over sampling and considering only the top 25 important variables.	12,000	1140	84.5	31.9

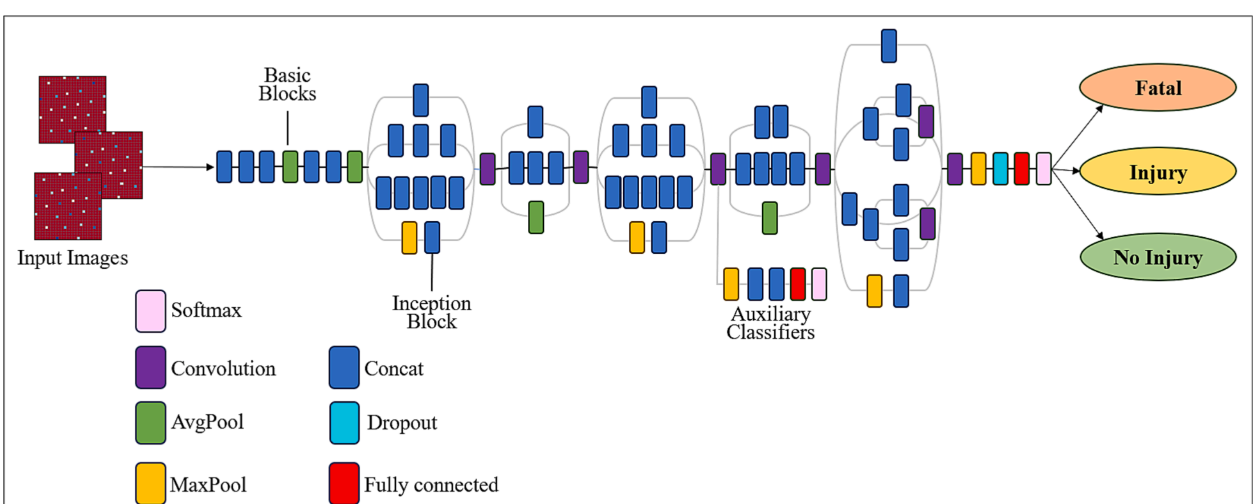

Fig. 2. Architecture of the inception-v3 Model for pedestrian crash severity prediction, adopted from (Meena et al., 2023).

batch size was minimized to 100 from 128. Moreover, the validation frequency was altered to 10 from 50. Subsequently, five distinct pedestrian crash models were constructed: MIM, employing the original imbalanced dataset; MBO, which used under sampling for data balance; MBU, incorporating over-sampling to balance the data; MBB, using both under-sampling and over sampling for data balance; and BBR, employing both under sampling and over sampling for data balance while only considering the top 25 significant variables. Among these, MNB utilized 8213 original imbalanced data images, achieving a validation accuracy of 84.4 % in a training duration of 19.9 min, as listed in Table 3. MBU, using a balanced set of 2,202 under-sampled images, managed a lower validation accuracy of 66.4 % but boasted the shortest training time of 4.6 min. MBO, processing an over-sampled balanced dataset of 19,872 images, yielded the highest validation accuracy of 87.0 % but required the longest training time of 63.2 min. MBB used a combination of under and over sampling on a balanced 12,000 image dataset to achieve a validation accuracy of 86.1 % within a training duration of 31.9 min. BBR, like MBB but solely considering the 25 most significant variables, reached a slightly lower validation accuracy of 84.5 % but managed the same training time. In conclusion, while MBO achieved the best validation accuracy, it required more training time. MBU had the quickest training but the least accuracy. MBB and MBR showed comparable performance metrics, even though BBR considered fewer variables.

Fig. 3 provides a visual representation of the MBB model's training progress. For brevity, only the progression of this specific model is depicted. The model's initial validation accuracy was roughly 42 %, which gradually improved with each training cycle until it reached a final validation accuracy of 86 % at the conclusion of the training period. In terms of validation loss, it was observed that it started off at approximately 1.1 in the initial phase and gradually decreased to an eventual value of around 1.4 by the end of the last training cycle.

To explore the performance of the pre-trained CNN further, the study employed the K-fold cross-validation technique. This method involves dividing the dataset randomly into K sets or folds of approximately equal size, with each of the K folds serving as validation data while the remaining $K-1$ folds are used to train the machine learning model. In this investigation, a 5-fold cross-validation approach was selected, a choice commonly recommended by many researchers and widely adopted in the applied machine learning domain (James et al., 2017). Table 4 presents the accuracy results from both 5-fold cross-validation and holdout validation. Although the 5-fold cross-validation method yielded slightly better validation accuracies compared to holdout validation, no significant differences were observed. Similar to holdout validation, the highest overall 5-fold cross-validation accuracy, at 87.3 % for predicting crash severity, was found for the MBO model, while the lowest accuracy, at 66.7 %, was recorded for the MBU model.

5.2. Variable importance

The study utilized the Boruta feature selection algorithm to investigate the impact of various variables on predicting pedestrian-involved crash severity. Fig. 4 presents a visual representation of the variable importance plot based on the Boruta algorithm. On the y-axis, the variable importance is represented using z-scores, while the corresponding variables are shown on the x-axis. The boxplots in green represent the variables deemed important in predicting crash severity, while the orange boxplots represent unimportant variables. Additionally, there are blue boxplots representing shadow features. Out of the 40 variables analyzed, the Boruta algorithm identified road alignment and weather conditions as unimportant in predicting pedestrian crash severity. It is worth noting that adverse weather conditions are often associated with increased motor vehicle crashes. However, in the case of pedestrian-involved crashes, adverse weather may show the opposite trend due to the reduced presence of pedestrians on the roadway during such conditions.

The Boruta algorithm effectively manages situations in which

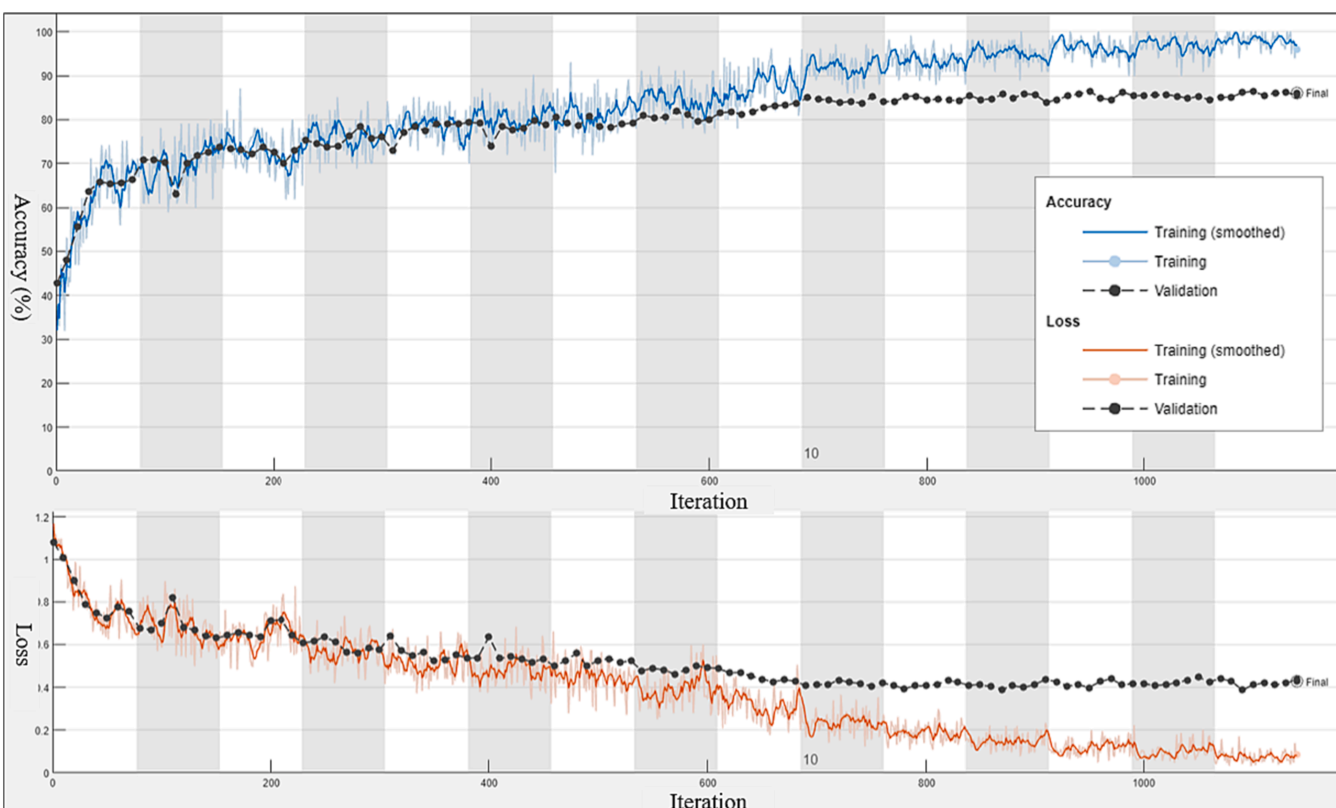

Fig. 3. Training progress of the pedestrian crash severity prediction model.

Table 4
Model Performance During Cross-Validation.

Model Name	5-Fold Cross-Validation						Holdout Validation Accuracy (%)
	Fold 1 Accuracy (%)	Fold 2 Accuracy (%)	Fold 3 Accuracy (%)	Fold 4 Accuracy (%)	Fold 5 Accuracy (%)	Overall Accuracy (%)	
MNB	85.5	85.2	84.4	84.6	84.9	84.9	84.4
MBU	66.6	67.3	66.2	66.8	66.5	66.7	66.4
MBO	87.9	87.1	86.5	87.6	87.4	87.3	87.0
MBB	87.5	86.1	86.2	87.2	86.9	86.8	86.1
MBR	85.7	84.8	85.2	85.4	84.9	85.2	84.5

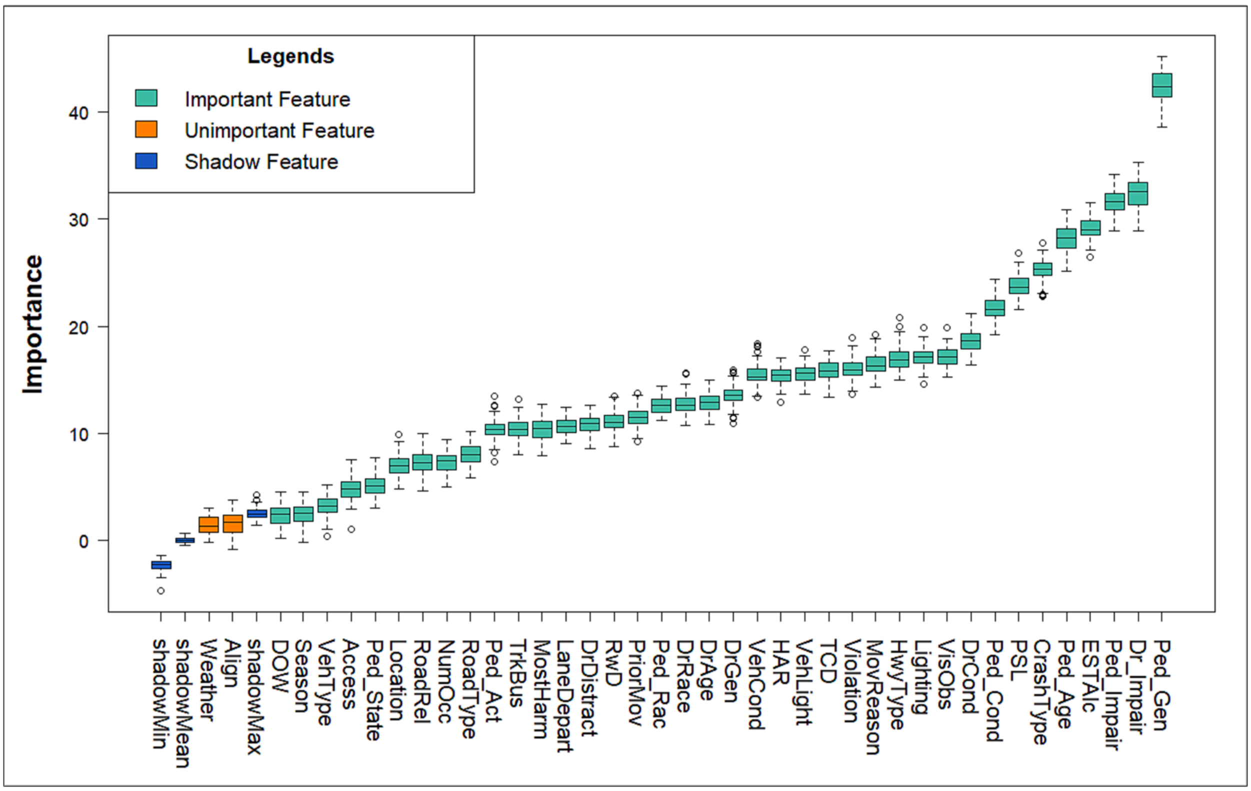

Fig. 4. Feature importance using boruta algorithm.

variables demonstrate similar levels of importance through its randomized approach and consensus mechanism. In instances where variables exhibit comparable importance levels, Boruta may yield diverse outcomes in multiple runs due to the inherent randomness in its process. To resolve these instances, Boruta incorporates a voting mechanism, taking into account the frequency with which variables are identified as important across multiple runs. The algorithm utilizes a Random Forest classifier to evaluate variable importance. It generates a shadow feature set by permuting values within each column, thereby creating a set of variables that represent noise or unimportant features. Subsequently, the algorithm assesses the importance of each variable in the original feature set by comparing it with the importance scores of the shadow features. Variables are then ranked based on their contribution to predictive accuracy. A Z-score is computed for each variable based on its relative importance, representing how many standard deviations the variable's importance deviates from the mean of the shadow features. Variables with Z-scores significantly higher than the mean are deemed important. Boruta conducts multiple iterations, each involving the creation of a new set of shadow features and a re-assessment of variable importance. The algorithm aggregates the results across iterations, considering variables that consistently emerge as important in the majority of runs as truly significant. Furthermore, Boruta considers variables that achieve a consensus of significance across iterations. If a variable is consistently identified as important, it is selected as a relevant feature. Conversely, variables with inconsistent importance across runs are considered unimportant. By employing this randomized and consensus-based approach, Boruta effectively addresses situations where variables display similar importance levels (Kursa et al., 2010; Kursa and Rudnicki, 2010). The algorithm's capacity to deliver stable and consistent results across multiple runs enhances its reliability in variable selection, contributing to a more robust understanding of feature importance in the context of pedestrian-involved crash severity prediction.

In Fig. 4, the analysis indicates that the gender of the pedestrian emerges as the most important variable in predicting pedestrian-involved crash severity. Gender differences may influence risk perception and decision-making while crossing the road or interacting with traffic. Moreover, physical differences between genders could also contribute to varying outcomes in pedestrian-involved crashes. Additionally, it's crucial to consider vulnerability and risk exposure, which can differ based on the type of pedestrian involved in the crash. Recent research has highlighted the impact of gender on pedestrian crashes. For instance, a recent study (de Armenta-Ramirez et al., 2023) found that older females are more susceptible to injuries during weekdays in downtown areas, while injured males are more prevalent during the afternoon in arterial streets. Understanding the influence of gender on pedestrian safety can inform targeted interventions and policies aimed at reducing crash severity.

Driver impairment, which refers to cases where drivers are impaired by alcohol, drugs, or both, emerged as the second most important factor associated with pedestrian crash severity. Impaired drivers experience reduced cognitive and motor skills, leading to diminished attention and delayed reactions, making it harder for them to detect and respond to pedestrians effectively. Additionally, impaired perception, characterized by altered depth perception and impaired judgment of distances, increases the likelihood of misjudging pedestrian positions and movements, elevating the risk of collisions. Furthermore, impaired drivers tend to engage in risky driving behaviors, such as speeding and disregarding traffic signals, which amplifies the danger to pedestrians. Numerous studies have identified driver impairment as one of the major contributing factors in pedestrian-involved crashes (Batouli et al., 2020, pp. 2006–2016; Yang et al., 2023).

The next important variable associated with pedestrian crash severity is related to pedestrian impairment, as depicted in Fig. 4. Impaired pedestrians may have diminished awareness of their surroundings and reduced attention to traffic, making them less likely to notice approaching vehicles or judge safe crossing opportunities accurately. Impaired pedestrians may have slower reaction times, hindering their ability to respond quickly to changing traffic conditions or unexpected situations. Previous study has also found that impaired pedestrians are more likely to be involved in crashes. For instance, a recent study (Hossain et al., 2023) reported that alcohol-impaired older pedestrians are more likely to be involved in fatal crashes. Another study (Batouli et al., 2020, pp. 2006–2016) also pedestrian impairment as a significant factors associated with the severity of pedestrian outcomes from motor vehicle crashes. Additionally, alcohol involvement, either by the driver or pedestrian (ESTAlc), emerged as another crucial variable. The variable importance analysis reveals that among the top five variables, three are related to impairment. This underscores the critical role of addressing impairment, both among drivers and pedestrians, as a key strategy to improve pedestrian safety and reduce the severity of pedestrian-involved crashes. Interventions such as public awareness campaigns, law enforcement efforts, and targeted educational programs can play a vital role in mitigating the risks associated with impairment and enhancing overall road safety for pedestrians and drivers alike.

Another essential variable related to pedestrian crash severity is pedestrian age. Previous studies have consistently shown that older pedestrians are more likely to be involved in severe crashes (de Armenta-Ramirez et al., 2023; Guo et al., 2023; Hossain et al., 2023). The posted speed limit is another critical factor associated with pedestrian crashes, supported by several previous studies (Adanu et al., 2023; Hossain et al., 2023; Islam, 2023). Higher speeds increase the impact force with the pedestrian, making them more susceptible to severe outcomes. Furthermore, an intriguing variable linked to pedestrian crash severity is visibility obstruction (VisObs). This refers to various obstacles, such as roadside objects, trees, or other vehicles, that block the line of sight for both drivers and pedestrians. When drivers' visibility is compromised, they may struggle to detect pedestrians in a timely manner, significantly increasing the risk of collisions. Urban environments, with their tall buildings and various structures, often present areas with limited visibility for both drivers and pedestrians, posing potential hazards for pedestrians. Other important variables related to pedestrian crashes include the type of crash, lighting conditions of roadways, and the conditions of both the pedestrian and the driver, such as distraction and inattentiveness, which have also been found to have a significant association with pedestrian crash severity.

5.3. Performance evaluation of pedestrian crash severity prediction model

Following the training and validation processes of the pedestrian-involved crash severity prediction models, their proficiency in predicting the severity of was assessed utilizing a separate test dataset comprising 20 % of the initial data. The prediction model's performance was examined through various performance metrics, such as accuracy, sensitivity, precision, specificity, F1-score, false negative ratio (FNR), and false positive ratio (FPR). These performance metrics were calculated using Equation (1) to Equation (7).

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

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Specificity = \frac{TN}{TN + FP} \tag{4}$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{5}$$

$$FNR = 1 - Recall \tag{6}$$

$$FPR = 1 - Specificity \tag{7}$$

where true positive, true negative, false positive, and false negative are represented by TP, TN, FP, and FN, respectively.

Five prediction models for pedestrian-involved crash severity were developed using the Inception-v3 deep learning model, taking into account various data balancing techniques. The performance metrics for these models are presented in Table 5. While a high level of overall accuracy is typically sought for deep learning-based prediction models, it doesn't necessarily indicate effective prediction performance when dealing with highly imbalanced data, a scenario often encountered in traffic crash severity data. Specifically, an elevated recall for the majority class, the injury crash group in this study, could boost the overall accuracy even when recall rates for minority classes, such as the fatal and no injury groups, are relatively low. The MNB model, which utilized the imbalanced original crash records, attained an overall accuracy of 82.8 %. Yet, when examining the accuracy of each crash severity level, i.e., the recall value, it is noticeable that only 51 % of fatal crashes were correctly classified, indicating a recall value of merely 51 % for this category. The no injury class demonstrated an even lower recall value of 28.7 %. The relatively low accuracies for the fatal and no injury categories can be attributed to the fact that these classes fall under the minority category, which has considerably fewer samples compared to the majority class (i.e., injury class). The recall value of the injury class was found to be 93.3 %, contributing significantly to the high overall accuracy.

In order to achieve balanced recall across all groups, this study initially implemented under sampling techniques using RUS, which resulted in 734 crash instances for each category. The model (MBU) built on this under-sampled dataset yielded an overall accuracy of 62.4 %, with recall values for fatal, injury, and no injury cases at 83.7 %, 52.4 %, and 51.0 %, respectively, as listed in Table 5. The lower prediction performance can be attributed to the fact that under-sampling drastically reduces crash instances from the majority group. This could lead to a significant loss of potentially useful information, as numerous valuable instances from the majority class are excluded from model training. Following this, the study employed over-sampling techniques based on the SMOTE, augmenting crash instances to 6,624 in each crash severity class. The model (MBO) generated from this data demonstrated considerably improved performance, with an overall accuracy of 87.6 % in predicting pedestrian-involved crash severity. The MBO model also balanced recall values across all categories, correctly predicting 94.6 %, 82.0 %, and 86.2 % of fatal, injury, and no injury instances, respectively.

Table 5
Prediction Performance of Pedestrian Involved-Crash Severity Models Using Inception-v3.

Model Name	Data Balancing	Number of images	Feature Selection	Category	Recall (%)	Precision (%)	Specificity (%)	FPR (%)	FNR (%)	F1-Score (%)	Accuracy (%)
MNB	No Balancing	Fatal = 734	All	Fatal	51.0	55.1	95.9	4.1	49.0	53.0	82.8
		Injury = 6,624		Injury	93.3	86.7	40.3	59.7	6.7	89.9	
		No Injury = 855		No Injury	28.7	60.5	97.8	2.2	71.3	38.9	
MBU	Under Sampling	Fatal = 734	All	Fatal	83.7	74.5	85.7	14.3	16.3	78.8	62.4
		Injury = 734		Injury	52.4	51.3	75.2	24.8	47.6	51.9	
		No Injury = 734		No Injury	51.0	59.5	82.7	17.3	49.0	54.9	
MBO	Over Sampling	Fatal = 6,624	All	Fatal	94.6	90.4	95.0	5.0	5.4	92.5	87.6
		Injury = 6,624		Injury	82.0	84.2	92.3	7.7	18.0	83.1	
		No Injury = 855		No Injury	86.2	88.0	94.2	5.8	13.8	87.1	
MBB	Both Under and Over Sampling	Fatal = 4,000	All	Fatal	93.5	88.5	93.9	6.1	6.5	90.9	85.6
		Injury = 4,000		Injury	77.5	82.7	91.9	8.1	22.5	80.0	
		No Injury = 4,000		No Injury	85.9	85.3	92.6	7.4	14.1	85.6	
MBR	Both Under and Over Sampling	Fatal = 4,000	Top 25 Important Variables	Fatal	93.0	87.6	93.4	6.6	7.0	90.2	84.6
		Injury = 4,000		Injury	74.3	81.5	91.6	8.4	25.8	77.7	
		No Injury = 4,000		No Injury	86.6	84.3	91.9	8.1	13.4	85.5	

Despite its benefits, over sampling can introduce overfitting to the model, where the model performs well on the training data but struggles when presented with new unseen crash instances that are slightly different.

To address the limitations of using solely over sampling or under sampling, this study concurrently employed both techniques, leading to a total of 12,000 crash instances, with each category comprising 4,000 samples. The model (MBB) using this data also yielded superior prediction performance, with an overall accuracy of 85.6 %. The recall and FNR values for the fatal categories were determined to be 93.5 % and 6.5 %, respectively, signifying that only 6.5 % of the fatal instances were misclassified. The prediction accuracy for the injury class was also found to surpass that of the other models. The data showed that only 22.5 % of the injury class cases were incorrectly predicted. Likewise, in the no-injury category, only 14.1 % of the crash instances were misclassified.

Finally, models for predicting the severity of pedestrian-involved crashes were formulated by taking into account the 25 most critical features identified by the Boruta algorithm, in conjunction with balanced data achieved through both over-sampling and under-sampling. The primary objective of this approach was not necessarily to enhance the model's accuracy, but to develop a model with a fewer number of variables. This approach is beneficial in reducing computational demand and improving the interpretability for transportation professionals. It is worth highlighting that in practical scenarios, pedestrian crash data and influencing factors might not always be easily accessible. As such, it is crucial to develop pedestrian-involved crash severity prediction models that rely on a limited number of features. The findings indicate that the model (MBR) with reduced variables also yielded high prediction performance. The overall accuracy of the MBR model was observed to be 84.6 %, which is slightly lower than the MBB model, as shown in Table 5. The recall values demonstrated that 93.0 % of fatal crashes, 74.3 % of injury crashes, and 86.6 % of no injury crashes were predicted accurately.

Fig. 5 displays the confusion matrices associated with the crash severity prediction models. The vertical axis represents the actual class while the horizontal axis indicates the predicted class. The digits along the diagonal line of the confusion matrix correspond to correct predictions for each respective crash severity category. Boxes adjacent to the confusion matrix and in line with the vertical axis exhibit recall values for the related categories. Similarly, boxes close to the confusion matrix and parallel to the horizontal axis present the precision values for the linked categories. The gradient of color within the boxes is indicative of the number of accurate predictions with darker shades signify a higher count of correct predictions. Fig. 5a presents the confusion matrix for the MNB model, which was trained without any data balancing. Although the overall accuracy was relatively high, the model's performance in predicting fatal and no injury crashes was suboptimal. More specifically, out of 147 tested fatal crashes, only 75 were correctly predicted, and out of 171 no-injury crashes, a mere 49 were accurately identified. The model based on under sampling (MBU) improved prediction rates, with significant enhancement noticeable only for the fatal class. Out of 147 tested fatal crashes, the MBU model accurately predicted 123, as demonstrated in Fig. 5b. The model relying on over sampling data (MBO) showcased the best performance. Out of 1325 test samples in each category, the MBO model correctly predicted 1,254 fatal crashes (94.6 %), 1,087 injury crashes (82.0 %), and 1,142 no-injury crashes (86.2 %), as illustrated in Fig. 5c. The model utilizing balanced data from both methods (MBB) also produced commendable prediction performance. For instance, out of 800 tested fatal crashes, this model accurately predicted 748 crashes (93.5 %), and misclassified 46 crashes to injury and 10 fatal crashes to no injury, as can be observed in Fig. 5d. Finally, the reduced model (MRB) also demonstrated comparable prediction performance. Out of 800 test samples in each category, the MRB model correctly predicted 744 fatal crashes, 594 injury crashes, and 693 no-injury crashes, as depicted in Fig. 5e.

5.4. Comparison of the Inception-v3 with the traditional Machine learning and statistical models

In order to evaluate the efficiency of the proposed Inception-v3 pretrained Convolutional Neural Network model in contrast to conventional machine learning models, six such traditional models were trained and examined in depth. These models include Classification and Regression Trees (CART), KNN, SVM, RF, Gradient Boosting (GB), and Naive Bayes (NB). For this comparative study, the balanced dataset utilized for the MBB model was used. To identify the best-performing models for each machine learning approach, the hyperparameters of these models were tuned instead of relying on their default values. Hyperparameters are specific properties that control various aspects of machine learning algorithms and can significantly impact their complexity and performance search. However, hyperparameter tuning is not a straightforward task, often relying on experimental results rather than theory. The most commonly used method for hyperparameter tuning is grid search (Chicco, 2017). In this study, the grid search

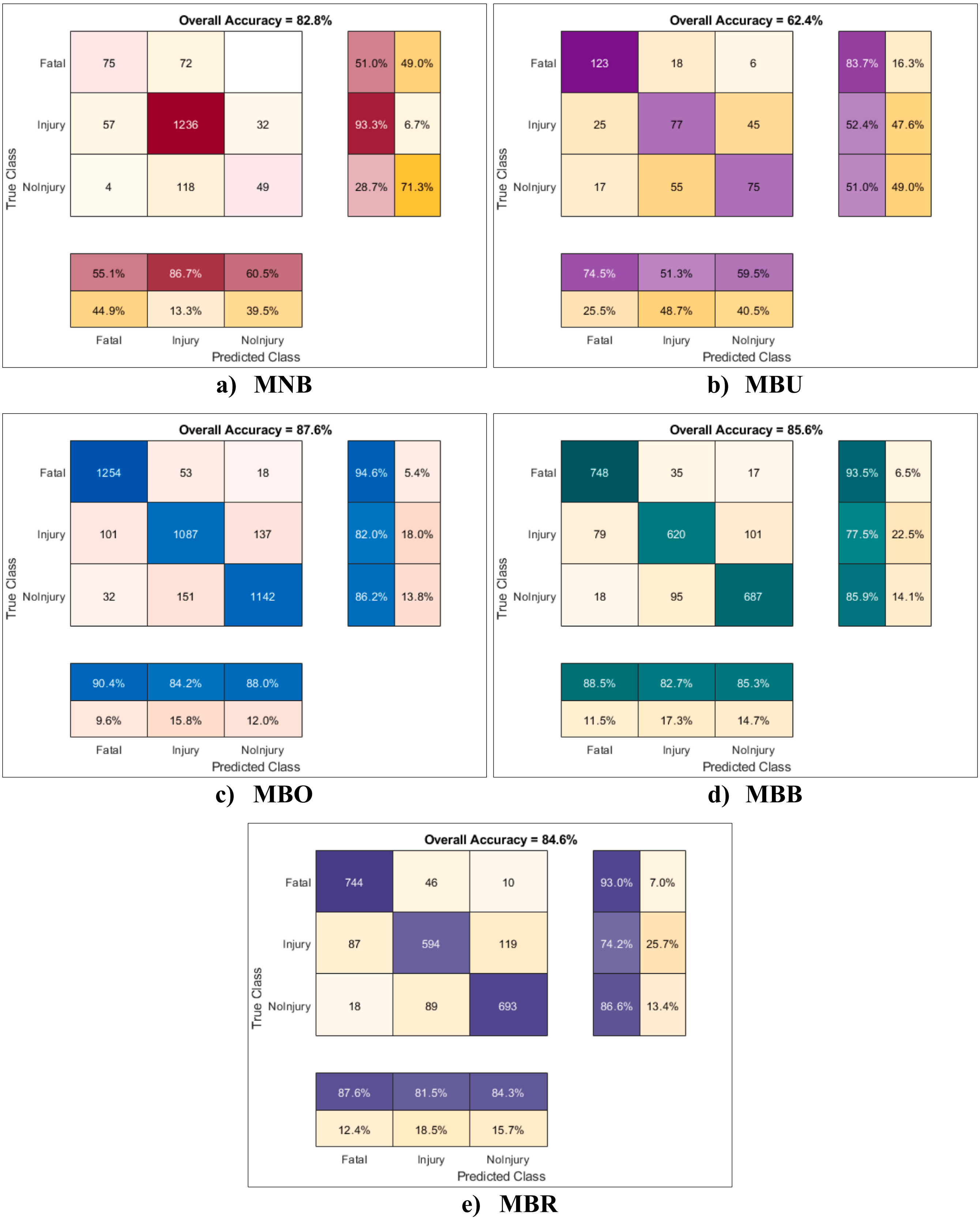

Fig. 5. Confusion matrix of the pedestrian-involved crash severity prediction models.

method was applied to find the best possible hyperparameters for each of the weather detection models using the “caret” package in R. Model performance was assessed using 5-fold cross-validation. The results of the parameter tuning for the crash prediction models are listed in Table 6, providing a detailed overview of the specific parameters and values used for each model.

After developing the best prediction models for each machine learning method, the prediction performance was compared with the Inception-v3 model, as reported in Table 7. The NB model was found to have the lowest performance, with an overall accuracy of 63.8 %. The CART model delivered an overall accuracy of 72.3 %. The KNN, SVM, and GB models demonstrated commendable overall prediction performances, with accuracy rates exceeding 77 %. However, the RF model produced the highest overall accuracy of 84.1 % among traditional machine learning models. In terms of recall or the model’s ability to predict pedestrian crashes across each severity level, the RF model also showcased the best results, accurately predicting 89.6 % of fatal crashes, 77.4 % of injury-inducing crashes, and 85.8 % of no injury crashes. Despite these findings, it is important to note that none of the traditional machine learning models managed to outperform the Inception-v3 in predicting pedestrian crash severity.

To further investigate the superiority of the Inception-v3 model over traditional machine learning models, McNemar’s Chi-squared test with continuity correction was leveraged. McNemar’s Test, is a statistical test tailored for paired nominal data, making it particularly relevant in the context of machine learning models (McNemar, 1947). This test can be used to compare the predictive accuracy of two models. The basis of McNemar’s Test is the use of a contingency table, designed to enumerate occurrences associated with two categorical variables. Within this framework, binary variables, correct/incorrect or yes/no, are examined across two distinct models, resulting in the creation of a 2 × 2 contingency table. The McNemar’s test statistic (chi-squared) is computed as follows:

$$\chi^2 = \frac{(x - y)^2}{x + y} \tag{8}$$

here, χ^2 represents McNemar’s test statistic, x denotes the number of correct classifications by the Inception-v3 model but misclassifications by a particular machine learning model using the same test dataset, and y represents the number of correct classifications by a particular machine learning model but misclassifications by the Inception-v3 model using the same test dataset. Edwards (1948) proposed a continuity-corrected version of the McNemar test to better align the test statistic with binary data. The corrected McNemar’s test statistic can be defined as follows:

Table 6
Parameter Tuning of the Machine Learning-Based Crash Severity Prediction Models.

Model	Parameters	Value
CART	Complexity parameter (cp)	6
KNN	Max number of neighbors (k)	5
SVM	Kernel function	radial
	Bandwidth of kernel function (Sigma)	0.016
	Cost (C)	0.5
RF	Number of randomly selected variables (mtry)	2
GB	Number of trees (n.trees)	150
	Maximum nodes per tree (interaction.depth)	3
	Learning rate (shrinkage)	0.1
	minimum number of observations in the terminal nodes (n. minobsinnode)	10
NB	Laplace correction (fL)	0
	Bandwidth of the kernel density (adjust)	1

Note: The terms inside parentheses represent the corresponding terminology used in the R packages for parameter configuration.

$$\chi^2 = \frac{(|x - y| - 1)^2}{x + y} \tag{9}$$

The null and alternative hypotheses of the McNemar’s test are as follows:

$$H_0 : p_I = p_M$$

$$H_1 : p_I \neq p_M \tag{10}$$

Here, the null hypothesis (H_0) suggests that there is no significant difference in prediction accuracy between the Inception-v3 model and the machine learning model, while the alternative hypothesis (H_1) suggests that a significant difference exists. As observed in Table 7, the p-values of McNemar’s test for all machine learning models are less than the significance level of 0.05. This indicates that the predictive performance of the Inception-v3 model is statistically superior compared to all machine learning models considered in this study. It is worth noting that while the overall accuracy of the RF model is comparable to the Inception-v3 model, the accuracy in predicting individual categories (i.e., recall values) reveals a disparity. Specifically, the RF model’s accuracy in correctly predicting fatal categories (89.6 %) is significantly lower compared to the Inception-v3 model (93.5 %). From a safety perspective, accurate and timely prediction of severe and injury crashes holds greater importance than predicting no-injury crashes. Furthermore, given that the deep learning model is based on images, there exists an opportunity to integrate machine vision-based data, including data gathered from sensors and video feeds, into the existing crash data. This integration can significantly enhance prediction accuracy, a capability that is not possible with the conventional machine learning-based model.

In addition to the machine learning models, this study also compared the Inception-v3 model with commonly used statistical models for categorical response variables. Three statistical models, namely multinomial logistic regression (MLR), ordered logit regression (OLR), and ordered probit regression (OPR), were developed and tested using both unbalanced and balanced data. The balanced data employed for these statistical models was the same dataset used to develop the MBR model, achieved through data balancing techniques, encompassing both under and oversampling methods. The performance of the statistical models is presented in Table 8. When the original unbalanced data was used, the MLR model achieved an overall accuracy of 83.7 %. However, upon a closer examination of individual category accuracy, specifically recall values for each crash severity category, it became evident that the MLR model excelled primarily in predicting the majority group, i.e., the injury category, with a notably high accuracy of 97.3 %. In contrast, it struggled to predict the minority groups, including the fatal and no injury categories, with only 40.5 % and 9 % correct predictions, respectively. The OLR and OPR models encountered even greater challenges, achieving only 17 % and 1.9 % correct predictions for the fatal and no injury categories using OLR, and 13.7 % and 1.3 % correct predictions for the fatal and no injury categories using OPR. To further evaluate the performance of the statistical models in predicting crash injury severity, these models were redeveloped using balanced data. Although the utilization of balanced data yielded more equitable results across the crash severity categories, the statistical models, when compared to the Inception-v3 model, still displayed significantly inferior performance. The MLR, OLR, and OPR models achieved overall accuracies of 67.9 %, 57.7 %, and 57.7 %, respectively, in contrast to the 85.6 % accuracy achieved by the Inception-v3 model.

These results highlight that while statistical models might excel in providing interpretability and understanding the marginal effects of various risk factors associated with crash occurrences, they fall short when it comes to developing highly accurate prediction models for real-world applications. In such scenarios, this study suggests the use of advanced deep learning models, like Inception-v3, due to their

Table 7
Comparison of Inception-v3 with Machine Learning Models.

Model	Category	Recall (%)	Precision (%)	Specificity (%)	FPR (%)	FNR (%)	F1-Score (%)	Accuracy (%)	McNemar's Test	
									chi-squared	p-value
CART	Fatal	80.0	80.0	89.7	10.3	20.0	80.0	72.3	142.8	< 0.001
	Injury	67.1	66.0	83.0	17.0	32.9	66.6			
	No Injury	69.6	70.8	85.7	14.3	30.4	70.2			
KNN	Fatal	98.3	82.3	89.2	10.8	1.7	89.6	80.5	32.2	< 0.001
	Injury	46.1	94.8	98.8	1.2	53.9	62.1			
	No Injury	96.5	73.6	82.8	17.2	3.5	83.5			
SVM	Fatal	85.5	85.5	92.6	7.4	14.5	85.5	77.8	60.3	< 0.001
	Injury	72.8	71.6	85.8	14.2	27.2	72.2			
	No Injury	74.9	76.1	88.3	11.7	25.1	75.5			
RF	Fatal	89.6	94.6	5.4	9.8	89.9	89.6	84.2	4.4	0.037
	Injury	77.4	88.3	11.7	18.0	79.7	77.4			
	No Injury	85.8	93.4	6.6	19.8	82.9	85.8			
GB	Fatal	84.3	84.9	92.3	7.7	15.7	84.6	79.1	44.5	< 0.001
	Injury	77.9	73.4	86.1	13.9	22.1	75.6			
	No Injury	75.1	79.3	90.3	9.7	24.9	77.2			
NB	Fatal	86.1	67.0	78.3	21.7	13.9	75.3	63.8	323.7	< 0.001
	Injury	27.3	75.0	95.5	4.5	72.7	40.0			
	No Injury	77.3	57.6	71.8	28.2	22.7	66.0			
Inception-v3	**Fatal**	**93.5**	**88.5**	**93.9**	**6.1**	**6.5**	**90.9**	**85.6**	–	–
	Injury	**77.5**	**82.7**	**91.9**	**8.1**	**22.5**	**80.0**			
	No Injury	**85.9**	**85.3**	**92.6**	**7.4**	**14.1**	**85.6**			

Table 8
Comparison of inception-v3 with statistical models.

Model	Data Balancing	Category	Recall (%)	Precision (%)	Specificity (%)	FPR (%)	FNR (%)	F1-Score (%)	Accuracy (%)
MLR	No Balancing	Fatal	40.5	66.0	97.9	2.1	59.5	50.2	83.7
		Injury	97.3	85.1	26.3	73.7	2.7	90.8	
		No Injury	9.0	60.9	99.4	0.6	91.0	15.7	
OLR		Fatal	17.0	54.2	98.5	1.5	83.0	25.9	81.6
		Injury	98.2	82.7	10.7	89.3	1.8	89.8	
		No Injury	1.9	33.3	99.6	0.4	98.1	3.7	
OPR		Fatal	13.7	61.8	99.1	0.9	86.3	22.5	81.9
		Injury	99.0	82.4	8.4	91.6	1.0	90.0	
		No Injury	1.3	40.0	99.8	0.2	98.7	2.5	
MLR	Both Under and Over Sampling	Fatal	77.2	79.8	90.0	10.0	22.8	78.5	67.9
		Injury	61.2	59.2	79.3	20.7	38.8	60.2	
		No Injury	65.1	65.1	82.7	17.3	34.9	65.1	
OLR		Fatal	72.1	77.7	89.4	10.6	27.9	74.8	57.7
		Injury	42.9	40.6	69.1	30.9	57.1	41.7	
		No Injury	57.8	56.8	78.2	21.8	42.2	57.3	
OPR		Fatal	72.3	77.9	89.5	10.5	27.7	75.0	57.7
		Injury	40.3	40.0	70.3	29.7	59.7	40.2	
		No Injury	60.1	56.4	76.9	23.1	39.9	58.2	
Inception-v3	Both Under and Over Sampling	**Fatal**	**93.5**	**88.5**	**93.9**	**6.1**	**6.5**	**90.9**	**85.6**
		Injury	**77.5**	**82.7**	**91.9**	**8.1**	**22.5**	**80.0**	
		No Injury	**85.9**	**85.3**	**92.6**	**7.4**	**14.1**	**85.6**	

significantly superior predictive capabilities, even at the expense of interpretability. It is important to note that the primary focus of this study was not to determine the superiority of one approach over the other but to illustrate how both methods can offer valuable insights to researchers and transportation practitioners. In other words, when interpretability is the primary focus, statistical models are more suitable, and when prediction accuracy is the primary objective, deep learning models are recommended.

6. Conclusion

In this research, a cutting-edge deep learning model, Inception-v3, was employed to construct a robust model for predicting the severity of pedestrian crashes. The study utilized five years of pedestrian crash severity records, spanning from 2016 to 2021, collected from the state of Louisiana. The final dataset incorporated forty variables pertaining to factors such as pedestrian characteristics, location, weather conditions, driver attributes, road features, and vehicle details. The severity of pedestrian crashes was categorized into three levels: fatal, injury, and no injury. To determine the importance of variables and to investigate the factors contributing to pedestrian crash severity, the Boruta algorithm was applied. Given the imbalanced nature of pedestrian crash data, this study adopted under-sampling using Random Under Sampling (RUS), and oversampling was performed using Synthetic Minority Over-sampling Technique (SMOTE) to balance the data. After balancing the data, the DeepInsight technique was utilized to convert the numeric crash data into image data. Subsequently, using the transformed images, five pedestrian crash severity prediction models were developed using the Inception-v3 model. These models were constructed considering various scenarios: 1) using the original data, 2) using the under-sampled data, 3) using the over-sampled data, 4) using a combination of under and over-sampled data, and 5) considering the top twenty-five important variables. Finally, the performance of the proposed pedestrian severity prediction model was evaluated and compared with several state-of-the-art machine learning models.

The Boruta feature selection analysis revealed several vital factors tied to pedestrian crash severity. These encompass pedestrian gender, pedestrian and driver impairment, posted speed limits, alcohol involvement, pedestrian age, visibility obstruction, roadway lighting conditions, and both pedestrian and driver conditions, including distraction and inattentiveness. The pedestrian crash severity prediction models unveiled that the model without any data balancing failed to deliver satisfactory detection performance, especially for minority classes such as the fatal and no injury categories. Remarkably, the application of under sampling techniques did not substantially improve the prediction performance, only yielding a 62.4 % overall accuracy. However, applying over sampling techniques substantially boosted prediction performance, resulting in an overall accuracy of 87.6 %, though at the cost of a significant increase in computational power requirements. Moreover, the model incorporating both over and under sampling techniques produced the most balanced performance in terms of overall accuracy and computational power. Specifically, this model achieved prediction accuracies of 93.5 %, 77.5 %, and 85.9 % in identifying fatal, injury, and no injury categories, respectively. The model based on the top twenty-five variables also demonstrated comparable performance, yielding an 84.6 % overall accuracy.

The research findings hold significant implications for diverse stakeholders, including safety professionals, emergency services, traffic control departments, and automobile manufacturers, providing valuable insights to enhance their safety measures. Beyond these applications, the findings can inform tailored pedestrian and driver safety education programs, addressing specific factors like impairment due to alcohol or distraction. Stricter speed limit enforcement in high pedestrian traffic areas, along with specialized educational initiatives for vulnerable age groups, can be implemented. Urban planning considerations may incorporate these insights, aiming to improve visibility and lighting conditions on roadways to reduce accidents. Additionally, the proposed approach, applicable to various safety contexts involving numerical or machine vision-based data, proves versatile. As connected and autonomous vehicles become more prevalent, the rise in telematics data collection, when integrated with images or videos, offers a comprehensive database. Applying the advanced Inception-v3 model to this comprehensive dataset enables a thorough analysis of pedestrian-involved crashes, fostering a deeper understanding and delivering vital insights for continual improvements in safety measures.

The proposed approach introduces an innovative model with exceptional predictive capabilities for assessing pedestrian-involved crash severity. Safety professionals and urban planners can leverage the model's strengths to avert potential crashes by predicting and addressing specific hotspots, particularly in urban landscapes susceptible to near-miss incidents. This research extends beyond theoretical advancements and could serve as a valuable tool for proactive safety measures. The model's efficacy in forecasting and mitigating risks in high-density pedestrian areas holds the promise of reshaping safety protocols. The practical application of the proposed methodology holds the potential to significantly transform safety practices, introducing a proactive and preventive approach. By integrating this methodology into safety protocols, urban planners and safety professionals can cultivate an environment where potential risks are identified and addressed before they escalate into critical incidents. This shift towards proactive safety measures is integral to ensuring the safety of pedestrians, especially in densely populated settings where the risk of crashes is inherently higher.

While this study provides significant insights into the prediction of pedestrian crash severity, it is not without limitations. First, the study relied on data exclusively from Louisiana, limiting the geographical applicability of the findings. Different regions may have unique factors that affect pedestrian crash severity. Future research should consider expanding the data set to include multiple regions to ensure more comprehensive and universally applicable results. For future research, it would be beneficial to extend the application of the Inception-v3 deep learning model to other road safety scenarios beyond pedestrian crashes.

CRediT authorship contribution statement

Md Nasim Khan: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **Subasish Das:** Conceptualization, Data curation; Formal analysis; Investigation; Methodology; Project administration; Resources; Supervision; Roles/Writing – original draft; and Writing – review & editing. **Jinli Liu:** Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Abdelwahab, H.T., Abdel-Aty, M.A., 2001. Development of Artificial Neural Network Models to Predict Driver Injury Severity in Traffic Accidents at Signalized Intersections: 10.3141/1746-02 1746 , 6–13. 10.3141/1746-02.

Adanu, E.K., Dzinyela, R., Agyemang, W., 2023. A comprehensive study of child pedestrian crash outcomes in Ghana. Accid. Anal. Prev. 189, 107146 https://doi.org/10.1016/j.aap.2023.107146.

Ahmadpour, H., Bazrafshan, O., Rafiei-Sardooi, E., Zamani, H., Panagopoulos, T., 2021. Gully erosion susceptibility assessment in the kondoran watershed using machine learning algorithms and the boruta feature selection. Sustainability (switzerland) 13 (18), 1–23. https://doi.org/10.3390/su131810110.

Alkheder, S., Taamneh, M., Taamneh, S., 2017. Severity Prediction of Traffic Accident Using an Artificial Neural Network. J. Forecast. 36 (1), 100–108. https://doi.org/10.1002/FOR.2425.

Armenta-Ramirez, I. de S., Reyes-Castro, P.A., Zuniga-Teran, A.A., Olmedo-Munoz, M., 2023. The urban structure and pedestrian injuries: A typological analysis of pedestrian crashes in the city of Hermosillo. Mexico. Traffic Injury Prevention 24 (5), 428–435. https://doi.org/10.1080/15389588.2023.2204386.

Batouli, G., Guo, M., Janson, B., Marshall, W., 2020. Analysis of pedestrian-vehicle crash injury severity factors in Colorado 2006–2016. Accid. Anal. Prev. 148, 105782 https://doi.org/10.1016/j.aap.2020.105782.

Bhat, C.R., Astroza, S., Lavieri, P.S., 2017. A new spatial and flexible multivariate random-coefficients model for the analysis of pedestrian injury counts by severity level. Anal. Methods Acc. Res. 16, 1–22. https://doi.org/10.1016/j.amar.2017.05.001.

Chawla, N.V., 2009. Data Mining for Imbalanced Datasets: An Overview. In: Maimon, O., Rokach, L. (Eds.), Data Mining and Knowledge Discovery Handbook. Springer, Boston, MA, pp. 875–886, 10.1007/978-0-387-09823-4_45.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: Synthetic Minority Over-sampling Technique. J. Artif. Intell. Res. 16, 321–357. https://doi.org/10.1613/JAIR.953.

Chicco, D., 2017. Ten quick tips for machine learning in computational biology. BioData Mining 10 (1), 1–17. https://doi.org/10.1186/s13040-017-0155-3.

Das, A., Khan, M.N., Ahmed, M.M., 2023. Deep learning approach for detecting lane change maneuvers using SHRP2 naturalistic driving data. Transport. Res. Rec. J. Transport. Res. Board 2677, 1.

Dong, Y., Zhang, Q., 2019. A combined deep learning model for the scene classification of high-resolution remote sensing image. IEEE Geosci. Remote Sens. Lett. 16 (10), 1540–1544.

Edwards, A.L., 1948. Note on the "correction for continuity" in testing the significance of the difference between correlated proportions. Psychometrika 13 (3), 185–187. https://doi.org/10.1007/BF02289261.

Ganganwar, V., 2012. An overview of classification algorithms for imbalanced datasets. Internat. J. Emerg. Technol. Adv. Eng. 2, 4.

Google, 2021. Imbalanced Data [WWW Document]. URL https://developers.google.com/machine-learning/data-prep/construct/sampling-splitting/imbalanced-data (accessed 7.24.21).

Goswamy, A., Abdel-Aty, M., Islam, Z., 2023. Factors affecting injury severity at pedestrian crossing locations with rectangular RAPID flashing Beacons (RRFB) using XGBoost and random parameters discrete outcome models. Accid. Anal. Prev. 181, 106937 https://doi.org/10.1016/j.aap.2022.106937.

Govinda, L., Sai Kiran Raju, M.R., Ravi Shankar, K.V.R., 2022. Pedestrian-vehicle interaction severity level assessment at uncontrolled intersections using machine learning algorithms. Saf. Sci. 153, 105806 https://doi.org/10.1016/j.ssci.2022.105806.

Guo, M., Yuan, Z., Janson, B., Yang, Y., 2020. A Two-Level Random Intercept Logit Model for Predicting Pedestrian-Vehicle Crash. p. pp 68-81. 10.1061/9780784483152.007.

Guo, M., Yuan, Z., Janson, B., Peng, Y., Yue, R., Zhang, G., 2023. Do factors associated with older pedestrian crash severity differ? A causal factor analysis based on exposure level of pedestrians. Traffic Inj. Prev. 24 (4), 321–330. https://doi.org/10.1080/15389588.2023.2183080.

Hossain, A., Sun, X., Thapa, R., Hossain, M.M., Das, S., 2023. Exploring association of contributing factors to pedestrian fatal and severe injury crashes under dark-no-streetlight condition. IATSS Research 47 (2), 214–224. https://doi.org/10.1016/j.iatssr.2023.03.002.

Hosseinzadeh, A., Moeinaddini, A., Ghasemzadeh, A., 2021. Investigating factors affecting severity of large truck-involved crashes: Comparison of the SVM and random parameter logit model. J. Saf. Res. 77, 151–160. https://doi.org/10.1016/j.jsr.2021.02.012.

Iranitalab, A., Khattak, A., 2017. Comparison of four statistical and machine learning methods for crash severity prediction. Accid. Anal. Prev. 108, 27–36. https://doi.org/10.1016/j.aap.2017.08.008.

Islam, M., 2023. An exploratory analysis of the effects of speed limits on pedestrian injury severities in vehicle-pedestrian crashes. J. Transp. Health 28, 101561. https://doi.org/10.1016/j.jth.2022.101561.

Jamal, A., Zahid, M., Tauhidur Rahman, M., Al-Ahmadi, H.M., Almoshaogeh, M., Farooq, D., Ahmad, M., 2021. Injury severity prediction of traffic crashes with ensemble machine learning techniques: a comparative study. Int. J. Inj. Contr. Saf. Promot. 28 (4), 408–427. https://doi.org/10.1080/17457300.2021.1928233.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2017. An Introduction to Statistical Learning: with Applications in R. Springer, New York.

Khan, M.N., Ahmed, M.M., 2020. Trajectory-level fog detection based on in-vehicle video camera with TensorFlow deep learning utilizing SHRP2 naturalistic driving data. Accid. Anal. Prev. 142, 105521 https://doi.org/10.1016/j.aap.2020.105521.

Kitali, A.E., Kidando, E., Sando, T., Moses, R., Ozguven, E.E., 2018. Predicting the Likelihood of Aging Pedestrian Severe Crashes Using Dirichlet Random-Effect Bayesian Logistic Regression Model. p. 17p.

Khan, M.N., Ahmed, M.M., 2023. A novel deep learning approach to predict crash severity in adverse weather on rural mountainous freeway. Journal of Transportation Safety & Security 15 (8), 795–825.

Khan, M.N., Das, A., Ahmed, M.M., 2024. Prediction of Truck-Involved Crash Severity on a Rural Mountainous Freeway Using Transfer Learning with ResNet-50 Deep Neural Network. Journal of Transportation Engineering, Part A: Systems 150 (2), 04023131.

Kong, X., Das, S., Zhang, Y., Wei, Z., Yuan, C., 2023. In-depth understanding of pedestrian-vehicle near-crash events at signalized intersections: an interpretable machine learning approach. Transp. Res. Rec. 2677 (5), 747–759. https://doi.org/10.1177/03611981221136138.

Kursa, M.B., Rudnicki, W.R., 2010. Feature selection with the Boruta package. J. Stat. Softw. 36, 1–13.

Kursa, M.B., Jankowski, A., Rudnicki, W.R., 2010. Boruta–a system for feature selection. Fund. Inform. 101 (4), 271–285.

Kwayu, K.M., Kwigizile, V., Oh, J.-S., 2020. Development of systemwide pedestrian safety performance function using stratified random sampling and a proxy measure of pedestrian exposure. Int. J. Inj. Contr. Saf. Promot. 27 (4), 420–431. https://doi.org/10.1080/17457300.2020.1791905.

Li, Y., Song, L., Fan, W., (David),, 2021. Day-of-the-week variations and temporal instability of factors influencing pedestrian injury severity in pedestrian-vehicle crashes: a random parameters logit approach with heterogeneity in means and variances. Anal. Methods Acc. Res. 29, 100152 https://doi.org/10.1016/j.amar.2020.100152.

McNemar, Q., 1947. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika 12 (2), 153–157. https://doi.org/10.1007/BF02295996.

Meena, G., Mohbey, K.K., Kumar, S., Chawda, R.K., Gaikwad, S.V., 2023. Image-based sentiment analysis using InceptionV3 transfer learning approach. SN Computer Science 4 (3), 242.

Mokhtarimousavi, S., Anderson, J.C., Azizinamini, A., Hadi, M., 2020. Factors affecting injury severity in vehicle-pedestrian crashes: A day-of-week analysis using random parameter ordered response models and Artificial Neural Networks. Int. J. Transp. Sci. Technol. 9 (2), 100–115. https://doi.org/10.1016/j.ijtst.2020.01.001.

Morris, C., Yang, J.J., 2021. Effectiveness of resampling methods in coping with imbalanced crash data: Crash type analysis and predictive modeling. Accid. Anal. Prev. 159, 106240 https://doi.org/10.1016/j.aap.2021.106240.

NHTSA, 2021. Traffic Safety Fact (2021 Data): Pedestrian (No. DOT HS 813 458). National Highway Traffic Safety Administration, Washington, DC.

Olowosegun, A., Babajide, N., Akintola, A., Fountas, G., Fonzone, A., 2022. Analysis of pedestrian accident injury-severities at road junctions and crossings using an advanced random parameter modelling framework: the case of Scotland. Accid. Anal. Prev. 169 https://doi.org/10.1016/j.aap.2022.106610.

Rahim, M.A., Hassan, H.M., 2021. A deep learning based traffic crash severity prediction framework. Acc. Anal. Prevent. 154 https://doi.org/10.1016/j.aap.2021.106090.

Sam, S.M., Kamardin, K., Sjarif, N.N.A., Mohamed, N., 2019. Offline signature verification using deep learning convolutional neural network (CNN) architectures GoogLeNet inception-v1 and inception-v3. Proc. Comput. Sci. 161, 475–483.

Sharma, A., Vans, E., Shigemizu, D., Boroevich, K.A., Tsunoda, T., 2019. DeepInsight: A methodology to transform a non-image data to an image for convolution neural network architecture. Sci. Rep. 9 (1), 1–7. https://doi.org/10.1038/s41598-019-47765-6.

Song, L., Li, Y., Fan (David), W., Wu, P., 2020. Modeling pedestrian-injury severities in pedestrian-vehicle crashes considering spatiotemporal patterns: insights from different hierarchical bayesian random-effects models. Anal. Methods Acc. Res. 28 https://doi.org/10.1016/j.amar.2020.100137.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In: Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826.

Toran Pour, A., Moridpour, S., Tay, R., Rajabifard, A., 2017. Modelling pedestrian crash severity at mid-blocks. Transportmet. A: Transport Sci. 13 (3), 273–297. https://doi.org/10.1080/23249935.2016.1256355.

Ukkusuri, S., Hasan, S., Aziz, H.M.A., 2011. Random parameter model used to explain effects of built-environment characteristics on pedestrian crash frequency. Transport. Res. Rec. J. Transport. Res. Board 2237, 98–106. https://doi.org/10.3141/2237-11.

van der Maaten, L., Hinton, G., 2008. Visualizing data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605. https://doi.org/10.1007/s10479-011-0841-3.

Wang, J., Huang, H., Xu, P., Xie, S., Wong, S.C., 2020. Random parameter probit models to analyze pedestrian red-light violations and injury severity in pedestrian-motor vehicle crashes at signalized crossings. J. Transport. Safety Security 12 (6), 818–837. https://doi.org/10.1080/19439962.2018.1551257.

Wang, X., Kim, S.H., 2019. Prediction and factor identification for crash severity: Comparison of discrete choice and tree-based models. Transport. Res. Rec. J. Transport. Res. Board 2673 (9), 640–653. https://doi.org/10.1177/0361198119844456.

Wen, X., Xie, Y., Jiang, L., Pu, Z., Ge, T., 2021. Applications of machine learning methods in traffic crash severity modelling: current status and future directions. Transp. Rev. 41 (6), 855–879. https://doi.org/10.1080/01441647.2021.1954108.

WHO, 2023. Road traffic injuries [WWW Document]. URL https://www.who.int/health-topics/road-safety (accessed 8.3.23).

Wu, Z., Misra, A., Bao, S., 2023. Modeling pedestrian injury severity: A case study of using extreme gradient boosting vs random forest in feature selection. Transport. Res. Rec. J. Transport. Res. Board. https://doi.org/10.1177/03611981231170014.

Yang, T., Fan (David), W., Song, L., 2023. Modeling pedestrian injury severity in pedestrian-vehicle crashes considering different land use patterns: Mixed logit approach. Traffic Inj. Prev. 24 (2), 114–120. https://doi.org/10.1080/15389588.2022.2156789.

Zamani, A., Behnood, A., Davoodi, S.R., 2021. Temporal stability of pedestrian injury severity in pedestrian-vehicle crashes: New insights from random parameter logit model with heterogeneity in means and variances. Anal. Methods Acc. Res. 32 https://doi.org/10.1016/j.amar.2021.100184.

Zeng, Q., Wang, Q., Zhang, K., Wong, S.C., Xu, P., 2023. Analysis of the injury severity of motor vehicle–pedestrian crashes at urban intersections using spatiotemporal logistic regression models. Accid. Anal. Prev. 189, 107119 https://doi.org/10.1016/j.aap.2023.107119.

Zhang, J., Li, Z., Pu, Z., Xu, C., 2018. Comparing prediction performance for crash injury severity among various machine learning and statistical methods. IEEE Access 6, 60079–60087. https://doi.org/10.1109/ACCESS.2018.2874979.

Zhu, C., Brown, C.T., Dadashova, B., Ye, X., Sohrabi, S., Potts, I., 2023. Investigation on the driver-victim pairs in pedestrian and bicyclist crashes by latent class clustering and random forest algorithm. Accid. Anal. Prev. 182, 106964 https://doi.org/10.1016/j.aap.2023.106964.

Zhu, Z., Hu, Z., Dai, W., Chen, H., Lv, Z., 2022. Deep learning for autonomous vehicle and pedestrian interaction safety. Saf. Sci. 145, 105479 https://doi.org/10.1016/j.ssci.2021.105479.