Project 2320 | December 2023

SJSU SAN JOSÉ STATE UNIVERSITY

MTI MINETA TRANSPORTATION INSTITUTE

Deep Learning, Machine Learning, or Statistical Models for Weather-related Crash Severity Prediction

Abimbola Ogungbire, M.S.
Panick Kalambay, Ph.D.
Hardik Gajera, Ph.D.
Srinivas Pulugurtha, Ph.D., P.E., F.ASCE

MINETA TRANSPORTATION INSTITUTE

Founded in 1991, the Mineta Transportation Institute (MTI), an organized research and training unit in partnership with the Lucas College and Graduate School of Business at San José State University (SJSU), increases mobility for all by improving the safety, efficiency, accessibility, and convenience of our nation's transportation system. Through research, education, workforce development, and technology transfer, we help create a connected world. MTI leads the Mineta Consortium for Transportation Mobility (MCTM) and the Mineta Consortium for Equitable, Efficient, and Sustainable Transportation (MCEEST) funded by the U.S. Department of Transportation, the California State University Transportation Consortium (CSUTC) funded by the State of California through Senate Bill 1 and the Climate Change and Extreme Events Training and Research (CCEETR) Program funded by the Federal Railroad Administration. MTI focuses on three primary responsibilities:

Research

MTI conducts multi-disciplinary research focused on surface transportation that contributes to effective decision making. Research areas include: active transportation; planning and policy; security and counterterrorism; sustainable transportation and land use; transit and passenger rail; transportation engineering; transportation finance; transportation technology; and workforce and labor. MTI research publications undergo expert peer review to ensure the quality of the research.

Education and Workforce Development

To ensure the efficient movement of people and products, we must prepare a new cohort of transportation professionals who are ready to lead a more diverse, inclusive, and equitable transportation industry. To help achieve this, MTI sponsors a suite of workforce development and education opportunities. The Institute supports educational programs offered by the Lucas Graduate School of Business: a Master of Science in Transportation Management, plus graduate certificates that include High-Speed and Intercity Rail Management and Transportation Security Management. These flexible programs offer live online classes so that working transportation professionals can pursue an advanced degree regardless of their location.

Information and Technology Transfer

MTI utilizes a diverse array of dissemination methods and media to ensure research results reach those responsible for managing change. These methods include publication, seminars, workshops, websites, social media, webinars, and other technology transfer mechanisms. Additionally, MTI promotes the availability of completed research to professional organizations and works to integrate the research findings into the graduate education program. MTI's extensive collection of transportation-related publications is integrated into San José State University's world-class Martin Luther King, Jr. Library.

Report 23-32

Deep Learning, Machine Learning, or Statistical Models for Weather-related Crash Severity Prediction

Abimbola Ogungbire, M.S.

Panick Kalambay, Ph.D.

Hardik Gajera, Ph.D.

Srinivas Pulugurtha, Ph.D., P.E., F.ASCE

December 2023

A publication of the
Mineta Transportation Institute
Created by Congress in 1991
College of Business
San José State University
San José, CA 95192-0219

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. 2320 23-32	**2. Government Accession No.**	**3. Recipient's Catalog No.**

4. Title and Subtitle Deep Learning, Machine Learning, or Statistical Models for Weather-related Crash Severity Prediction	**5. Report Date** December 2023 **6. Performing Organization Code**
7. Authors Abimbola Ogungbire, M.S. ORCID: 0000-0002-1176-0422 Panick Kalambay, Ph.D. ORCID: 0000-0002-0815-3074 Hardik Gajera, Ph.D. ORCID: 0000-0001-9010-1355 Srinivas Pulugurtha, Ph.D., P.E., F.ASCE ORCID: 0000-0001-7392-7227	**8. Performing Organization Report** CA-MTI-2320
9. Performing Organization Name and Address Mineta Transportation Institute College of Business San José State University San José, CA 95192-0219	**10. Work Unit No.** **11. Contract or Grant No.** 69A3551747127
12. Sponsoring Agency Name and Address U.S. Department of Transportation Office of the Assistant Secretary for Research and Technology University Transportation Centers Program 1200 New Jersey Avenue, SE Washington, DC 20590	**13. Type of Report and Period Covered** **14. Sponsoring Agency Code**

15. Supplemental Notes

10.31979/mti.2023.2320

16. Abstract

Nearly 5,000 people are killed and more than 418,000 are injured in weather-related traffic incidents each year. Assessments of the effectiveness of statistical models applied to crash severity prediction compared to machine learning (ML) and deep learning techniques (DL) help researchers and practitioners know what models are most effective under specific conditions. Given the class imbalance in crash data, the synthetic minority over-sampling technique for nominal (SMOTE-N) data was employed to generate synthetic samples for the minority class. The ordered logit model (OLM) and the ordered probit model (OPM) were evaluated as statistical models, while random forest (RF) and XGBoost were evaluated as ML models. For DL, multi-layer perceptron (MLP) and TabNet were evaluated. The performance of these models varied across severity levels, with property damage only (PDO) predictions performing the best and severe injury predictions performing the worst. The TabNet model performed best in predicting severe injury and PDO crashes, while RF was the most effective in predicting moderate injury crashes. However, all models struggled with severe injury classification, indicating the potential need for model refinement and exploration of other techniques. Hence, the choice of model depends on the specific application and the relative costs of false negatives and false positives. This conclusion underscores the need for further research in this area to improve the prediction accuracy of severe and moderate injury incidents, ultimately improving available data that can be used to increase road safety.

17. Key Words Crash, Weather, Deep learning, Machine learning, Statistical analysis.	**18. Distribution Statement** No restrictions. This document is available to the public through The National Technical Information Service, Springfield, VA 22161.

19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 49	**22. Price**

Form DOT F 1700.7 (8-72)

DOI: 10.31979/mti.2023.2320

Mineta Transportation Institute
College of Business
San José State University
San José, CA 95192-0219
Tel: (408) 924-7560
Fax: (408) 924-7565
Email: mineta-institute@sjsu.edu

transweb.sjsu.edu/research/2320

ACKNOWLEDGMENTS

The authors sincerely thank the Highway Safety Information System (HSIS) for the help with the data required for this research. They also appreciate Dr. Sonu Matthew for assisting with problem formulation and preliminary data processing.

CONTENTS

List of Figures

List of Tables

Executive Summary

The focus of this study is to evaluate and compare the performance of three different approaches for crash severity prediction: statistical, machine learning (ML), and deep learning (DL) models. The goal was to find the most effective model for each crash severity level and understand the strengths and limitations of each approach.

The performance of the ordered probit model (OPM), a statistical model, is similar to the ordered logit model (OLM), but both struggle to correctly identify severe-injury crashes and distinguish between moderate-injury and property-damage-only (PDO) crashes. In the realm of ML, models such as random forest (RF) and XGBoost are evaluated. For DL, models such as the multi-layer perceptron (MLP) and TabNet are assessed. The performance of these models varied across severity levels, with PDO predictions being the best and severe-injury predictions performing the worst. In the case of severe-injury classification, the best performing model is reported to be TabNet with a precision of 8.45%, a recall of 45.73%, and an F1 score of 28.78%. The model with the best performance for predicting moderate-injury crashes is RF, with a precision of 42.77%, a recall of 85.53%, and an F1 score of 57.03%. For PDO classification, TabNet again performed best with a precision of 79.62%, a recall of 82.72%, and an F1 score of 92.10%. However, all these models struggled with severe-injury classification due to factors such as class imbalance or the complexity of the problem.

It can be concluded that the statistical models have high precision but low recall for severe-injury crashes, suggesting they are confident but not comprehensive in identifying severe-injury crashes. The ML models offer a more balanced performance with moderate precision and recall. In contrast, the DL models have high recall but low precision, indicating they may incorrectly classify incidents as severe- or moderate-injury crashes. The choice of a model may depend on the specific application and the relative costs of false negatives and false positives. For example, a model with a high recall (such as the DL models) might be preferable in preliminary screening tools where the cost of a false negative is much higher than the cost of a false positive.

1. Introduction

Road traffic crashes are a significant public health concern (Gopalakrishnan, 2012; CDC, 2023). Rain, snow, ice, and poor visibility often create unsafe road conditions for drivers (Druta et al., 2020; Greaves et al., 2023). Recent reports published by the Federal Highway Administration (FHWA) indicate that nearly 5,000 people are killed and more than 418,000 people are injured in weather-related crashes each year (FHWA, 2020). The consequences of a driver's failure to stop when confronted by either poor visible conditions or wet pavement frequently result in severe crashes. In order to reduce the intensity or frequency of these weather-related crashes, it is important to understand their contributing factors and to select and implement relevant countermeasures. In addition, providing an assessment of the effectiveness of statistical models, machine learning (ML) models and deep learning (DL) models applied to crash severity prediction will help researchers and practitioners know which models are most effective under specific conditions.

Researchers in the past adopted statistical methods such as binary logistic regression models, multinomial logit models, OPMs, and random parameter logit models to evaluate crash severity (Penmetsa & Pulugurtha, 2017, 2019; Penmetsa et al., 2018a; Chen et al., 2019; Tagar & Pulugurtha, 2021; Fanyu et al., 2021; Hou et al., 2022). Historically, statistical models have been widely utilized for crash severity prediction based on weather variables. Researchers employed regression techniques to establish linear relationships between weather conditions and crash outcomes (Mao et al., 2019; Abdulhafedh, 2022). Although these models provided initial insights, they often failed to capture intricate nonlinear interactions and complex patterns in the data. The emergence of ML techniques revolutionized crash severity prediction by enabling the development of more sophisticated models. Decision trees, random forests (RFs), and support vector machines (SVMs) were among the first algorithms employed in the context of crash severity analysis (Theofilatos et al., 2019; Hadjidimitriou et al., 2019; Mokhtarimousavi et al., 2020). These models introduced the concept of feature importance, enabling researchers to identify the most influential weather-related variables affecting crash severity. DL, a subset of ML, has gained substantial popularity in recent years due to its ability to automatically learn hierarchical representations from raw data. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have been leveraged to extract intricate patterns from weather data, empowering more accurate crash severity predictions. These models excel at handling complex temporal and spatial dependencies within weather variables, capturing subtle nuances that were previously overlooked. A few studies were conducted to evaluate the performance of statistical models, ML models, and DL models in weather-severity crash severity prediction. These studies often employed large datasets encompassing diverse weather conditions and incorporating a range of evaluation metrics such as accuracy, precision, recall, and F1 score. The results varied across these studies, with each modeling technique demonstrating its strengths and limitations.

A diverse set of models were chosen to comprehensively assess their effectiveness in predicting weather-related crash severity. OLM and OPM were chosen due to their established use in similar research and their interpretability, which is well-suited for ordinal crash severity data. ML techniques such as RF and XGBoost were chosen for their robustness in handling noisy data, capturing intricate interactions, and providing feature importance insights. These models strike a balance between accuracy and interpretability. In the realm of DL, the multi-layer perceptron neural network (MLP NN) and TabNet were chosen, leveraging their capacity to extract complex non-linear patterns from weather data, particularly adept at handling the temporal and spatial dependencies within weather variables. TabNet, designed for tabular data, offers feature selection capabilities. By including this array of models, the study aims to offer a comprehensive understanding of their strengths and limitations, thereby providing valuable insights for predicting crash severity under various weather conditions. These models are designed to forecast the level of traffic crash injuries, specifically factoring in prevailing weather conditions as one of their inputs. In other words, these models aim to estimate the severity of injuries resulting from a traffic crash, and they consider the weather conditions at the time of the crash as a significant factor in making these predictions.

1.1 Organization of the Report

The remainder of the report comprises four chapters. Chapter 2 presents and analyzes the literature of weather-related crashes, including the methods. Chapter 3 describes the data, study area, and method used for crash severity analyses in this study. Chapter 4 presents and analyzes the results from this study, and Chapter 5 summarizes the conclusions and scope for future work.

2. Literature Review

2.1 Overview of Weather-related Crashes

Weather-related crashes pose significant risk to road users and have been a subject of extensive research in the transportation safety domain. Numerous studies were conducted to investigate the relationship between weather conditions and crash occurrence (Zhao et al., 2019; Al-Mistarehi et al., 2022; Wei et al., 2022). Adverse weather factors such as rain, snow, fog, and icy roads have been found to significantly increase the likelihood of crashes (Hammad et al., 2019; Robins & Fotio, 2020; Das et al., 2020; Zhang et al., 2021). For instance, studies have reported higher crash rates during rainfall or heavy precipitation events (Das et al., 2020). Additionally, reduced visibility due to fog or mist has been linked to an increased crash risk (Robins & Fotio, 2020). Understanding the specific weather variables and their impact on crash occurrence is essential for developing targeted interventions.

Weather conditions play a critical role in determining crash severity. Several studies have explored the association between weather variables and the severity of crashes (Xing et al., 2019; Abohassan et al., 2021; Yang et al., 2022). Harsh weather conditions, including heavy rain, snowstorm, and high wind, have been linked to an increased likelihood of severe crashes (Xing et al., 2019). Poor road surface conditions during adverse weather, such as slippery roads or reduced tire grip, contribute to more severe crashes (Druta et al., 2020). Investigating the relationship between weather conditions and crash severity is crucial for allocating resources and implementing appropriate safety measures.

Studies on weather-related crashes utilize diverse data sources, including police reports, crash databases, weather station data, and advanced sensor technologies (Das et al., 2020; Duddu et al., 2020; Mathew & Pulugurtha, 2022). These datasets provide valuable information for conducting analyses. However, they may have limitations such as underreporting crashes or incomplete weather data. Addressing data limitations and improving data collection methods remain critical challenges in weather-related crash research. Understanding the impact of weather conditions on crash occurrence and severity has practical implications for road safety management.

While significant effort has been made in crash severity analysis, there is a clear need for more comprehensive studies that build on identifying the best method in analyzing weather-related crash severity. This study compares a broader range of statistical, ML, and DL models on the same representative crash dataset. This research could provide greater understanding of the intricate relationship between weather conditions and crash severity, and ultimately contribute to the development of more effective strategies for enhancing road safety during adverse weather conditions.

2.2 Previous Research on Crash Severity Prediction

Crash severity prediction is a critical area of research that aims to identify factors and develop models to accurately predict the severity of traffic crashes. Accurate crash severity prediction can aid in emergency response planning, resource allocation, and the development of effective countermeasures to reduce crash severity. Numerous studies were conducted to investigate the factors influencing crash severity (Zhao et al., 2019; Das et al., 2020; Al-Mistarehi et al., 2022; Wei et al., 2022; Yang et al., 2022). These factors can be broadly categorized into three main groups: driver-related, vehicle-related, and environmental-related factors (Robins & Fotio, 2020; Hou et al., 2022; Yang et al., 2022). Driver-related factors include driver age, gender, impairment (e.g., alcohol or drug use), distraction, and fatigue (Dingus et al., 2016). Vehicle-related factors encompass vehicle type, size, and safety features. Environmental-related factors consist of road conditions, weather conditions, lighting, and traffic characteristics. Understanding the impact of these factors is crucial for developing effective crash-severity prediction models. The selection of predictor variables significantly impacts the accuracy and interpretability of crash-severity prediction models (Sattar et al., 2023). Previous research has identified a wide range of potential predictor variables, including driver characteristics (e.g., age and gender), roadway attributes (e.g., speed limit and road type), environmental conditions (e.g., weather and lighting), and crash-specific variables (e.g., crash type and time of day) (Duddu et al., 2019; Shi et al, 2019; Yuan et al., 2019; Islam & Mannering, 2020). Feature selection techniques such as stepwise regression, principal component analysis, or recursive feature elimination have been employed to identify the most influential variables for crash-severity prediction.

Evaluation metrics play a crucial role in assessing the performance of crash-severity prediction models. Commonly used metrics include accuracy, precision, recall, F1 score, and area under the curve—receiver operating characteristic curve (AUC-ROC) (Yuan et al., 2019). Additionally, confusion matrix analysis provides insights into model performance across different severity levels. Studies have compared the performance of different models and techniques, highlighting the strengths and limitations of each approach. Higher accuracy and AUC-ROC values indicate better model performance (Ke et al., 2017). Accurate crash-severity prediction models have practical implications for road safety management. For instance, emergency response systems can use predicted severity levels to dispatch appropriate medical personnel and resources. Transportation agencies can prioritize road safety improvements and allocate funding based on predicted crash severity hotspots (Jamal et al., 2021). Furthermore, crash-severity-prediction models can aid in the development of intelligent transportation systems and advanced driver-assistance systems to prevent or mitigate crashes.

2.3 Methods in Crash Severity Analysis

Researchers have widely used statistical models in the past to determine the effect of factors affecting crashes on injury severity levels. Amongst statistical models, discrete choice models are widely employed in the existing literature considering the discrete nature of variables in crash datasets (Penmetsa & Pulugurtha, 2018b; Yan et al., 2021). Logistic regression techniques are widely used in cases when the dependent variable is discrete or ordinal. Several researchers used different models such as the OLM (Ayuso and Santolino, 2007), OPM (Abdel-Aty 2003; Gray et al., 2008; Garrido et al., 2014; Penmetsa et al., 2017), Bayesian OPM (Xie et al., 2009), probit or logit models with random parameters (Behnood and Mannering 2015; Behnood and Mannering 2016; Behnood and Mannering 2017), hierarchical OLMs, and OPMs (Chen et al., 2016; Fountas and Anastasopoulos, 2018). Advanced statistical models provide flexibility to account for heterogeneity due to unobserved variables as well as account for random or hierarchical components (Duddu et al., 2018; Duvvuri et al., 2022).

Traditional OLMs and OPMs are used in this study since they are computationally efficient and do not account for any random component in modeling. Traditional models also provide flexibility to compare the accuracy matrix with an ML or DL model, with results employed on the same dataset.

With the advancement of computational power and data availability, ML models have started to gain attention in traffic crash-severity prediction. These models include decision trees, RFs, SVMs, and gradient boosting machines, among others. Zeng et al. (2019) employed a DL approach, specifically a long short-term memory (LSTM) model, to predict crash severity. Their study demonstrated that LSTM, a type of recurrent neural network, is well suited to handling time-series data and can effectively model the temporal dependencies of various factors contributing to crash severity. Chen et al. (2016) used a decision tree model and found it to outperform logistic regression in terms of prediction accuracy. However, the model might overfit the data if not properly tuned. Wang et al. (2019) showed that SVMs, combined with an effective feature selection strategy, provided highly accurate predictions and outperformed traditional logistic regression models. Santos et al. (2020) compared various ML models, including decision trees, RFs, and XGBoost, for crash severity prediction and found XGBoost to deliver the best performance. Tang et al. (2020) reviewed the use of RF models for crash severity prediction and highlighted their superior performance over traditional statistical models, particularly in handling high-dimensional and non-linear data. Nevertheless, the interpretability of RF models remains a challenge.

Deep learning (DL) models for crash severity prediction have been used in a limited number of studies (Abdelwahab & Abdel-Aty, 2001; Alkheder et al., 2017; Das et al., 2018; Zheng et al., 2019; Ma et al., 2021; Rahim & Hassan, 2021; Khan & Ahmed, 2022). Abdelwahab & Abdel-Aty (2001) employed multi-layer perceptron (MLP) and fuzzy adaptive resonance theory (ART) to develop driver-injury severity models. Their DL models outperformed OLMs, achieving an accuracy of 65.6% and 60.14% on the training and testing datasets, respectively. Alkheder et al. (2017) also demonstrated the superiority of DL models, specifically artificial neural networks (ANNs), over the OPM. Their ANN-based model achieved an accuracy of 81.6% and 74.6% on the training and testing datasets, respectively. Das et al. (2018) introduced "DeepScooter," a deep learning framework based on MLP, for predicting crash severities involving at-fault motorcycle riders. Remarkably, their framework achieved unprecedented accuracies of 100% and 94% on the training and testing datasets, respectively. Ma et al. (2021) proposed an analytic framework utilizing a DL model. Their model achieved a recall value of 0.82–0.85 for geographical clusters of fatal and serious injury crashes. In contrast to the aforementioned studies, Zheng et al. (2019), Rahim & Hassan (2021), and Khan & Ahmed (2022) utilized CNN-based models to predict crash severity levels. Zheng et al. (2019) developed a novel CNN-based model named "TASP-CNN" for crash-severity prediction. Their model achieved a recall of 6.3% on fatal crashes (16.7% and 93.2% for slight and serious injury crashes, respectively) and an overall F1 score of 87%. Rahim & Hassan (2021) proposed a CNN-based framework for predicting crash severity in highway work zones, utilizing the EfficientNet model (Tan et al., 2019). They achieved the highest recall of 67% for fatal crashes. More recently, Khan & Ahmed (2022) developed a crash-severity-prediction model for rural mountainous freeways, employing the ResNet18 algorithm. This model achieved a recall of 99.3% for fatal crashes. It is worth noting that, unlike the other mentioned studies, their study focused solely on predicting crash severity in adverse weather conditions.

Despite the extensive literature on traffic crash-severity prediction using statistical, ML, and DL models, there is still a lack of studies that thoroughly compare these models in the same context, using the same dataset. Therefore, a comprehensive comparison study is necessary. This study aims to fill these gaps by conducting a comprehensive comparison of different models in terms of various performance metrics. This study will provide a more complete picture of the strengths and weaknesses of different models- and potentially contribute to the development of more effective and efficient traffic safety strategies.

3. Methodology

3.1 Data Source

The data used in this study were obtained from the Highway Safety Information System (HSIS) database. Crashes that took place between January 1, 2015 and December 31, 2017 were extracted. Crashes are reported using case numbers and observations with the same number indicating that the vehicles involved are part of the same crash incident. To gain a thorough understanding of the crash occurrence process, Washington and Haque (2013) argued that crashes due to different causes should be modeled separately. Hence, only weather-related crashes were extracted, while crashes occurring under clear weather conditions were excluded to obtain a final dataset. The final dataset only included crashes that happened under non-clear weather conditions (i.e., with cloudy, rain, fog/smog, sleet/hail/freezing rain/drizzle, severe crosswinds, or blowing sand conditions described in the crash reports).

Crash severity is defined in the HSIS database at five different levels: fatal crashes, injury type class A, injury type class B, injury type class C, and no injury/property damage only (PDO). For this analysis, the crash severity was re-categorized into three levels, i.e., severe injury (fatal and injury type class A), moderate injury (injury type class B and injury type class C), and PDO/no injury. A total of 238,252 weather-related crashes were recorded within the study period with 2,952 severe crashes, 71,688 moderate crashes, and 163,612 PDO crashes.

Table 1 presents the summary statistics providing information about the various categories and variables associated with weather-related crashes. It shows the count and percentage of severe, moderate, and PDO crashes for each category, such as weather condition, contributing factor of the crash, road surface condition, functional class of the road, location type, light condition, road characteristic, driver gender, driver age, speed limit class, crash type, work zone area, vehicle type, seasonal factors, road terrain, time of day, day of the week, locality, and more.

The summary statistics give information about the counts and percentages of different types of weather conditions for severe, moderate, and PDO injury cases. Cloudy weather is the most common weather condition in all three categories, with 63.4% of severe crashes, 57.6% of moderate-injury crashes, and 56.6% of no-injury crashes occurring under cloudy conditions. Rain is the second most common weather condition, followed by snow, fog, smog, and smoke. Sleet, hail, and freezing rain/drizzle are less common, with blowing sand and dirt being the least common weather condition for all three severity types.

Table 1. Summary Statistics

Variable	Category	Description	Severe Injury		Moderate Injury		PDO	
			Count	%	Count	%	Count	%
Weather condition	1	Cloudy	1871	63.4	41293	57.6	92663	56.6
	2	Rain	893	30.3	26828	37.4	60374	36.9
	3	Snow	41	1.4	1085	1.5	4296	2.6
	4	Fog, smog, smoke	104	3.5	1247	1.7	2786	1.7
	5	Sleet, hall, freezing rain/drizzle	39	1.3	1196	1.7	3324	2
	6	Severe crosswinds	3	0.1	31	0.04	132	0.08
	7	Blowing sand, dirt	1	0.03	8	0.01	37	0.02
Contributing factor of the crash	1	No contributing factors	1191	40.3	32095	44.8	75111	45.9
	2	Disregarding signs or signals	98	3.3	2124	3.0	2492	1.5
	3	Exceeded safe speed/speed limit or failed to reduce speed	548	18.6	17302	24.1	41651	25.5
	4	Improper turn or right turn on red	17	0.6	730	1.0	1979	1.2
	5	Crossed centerline, improper lane change, or use of an improper lane	222	7.5	2135	3.0	6102	3.7
	6	Overcorrected, oversteered, improper passing, or improper backing	77	2.6	1503	2.1	3019	1.8
	7	Failing to yield to the right-of-way, or driver inattention	267	9.0	9147	12.8	18870	11.5
	8	Operating too closely, aggressive driving, or alcohol use	429	14.5	3768	5.3	6578	4.0
	9	Visibility obstruction, or defective equipment	16	0.5	471	0.7	1275	0.8
	10	Other/unable to determine	87	2.9	2413	3.4	6535	4.0
Road surface condition	1	Dry	1290	43.7	27329	38.1	60413	36.9
	2	Wet, presence of water (standing/moving)	1577	53.4	41617	58.1	93882	57.4
	3	Ice, snow, slush	81	2.7	2708	3.8	9248	5.7
	4	Sand, mud, dirt, gravel, fuel, or oil	4	0.1	34	0.05	69	0.04
Functional class of road	1	Principal arterial: interstate, freeways and expressways	402	13.6	12617	17.6	35002	21.4
	2	Principal arterial: other	682	23.1	22955	32.0	50092	30.6
	3	Minor arterial	693	23.5	18303	25.5	39919	24.4
	4	Major collector	789	26.7	11795	16.5	24314	14.9
	5	Local	386	13.1	6018	8.4	14285	8.7
Location type	0	Non-intersection	2516	85.2	57486	80.2	138677	84.8
	1	Intersection	436	14.8	14202	19.8	24935	15.2
Light condition	1	Daylight	1727	58.5	51558	71.9	118084	72.2
	2	Dusk, and dawn	163	5.5	3582	5.0	7799	4.8
	3	Dark lighted roadway/unknown lighting	244	8.3	7146	10.0	15206	9.3
	4	Roadway not lighted	818	27.7	9402	13.1	22523	13.8

Variable	Category	Description	Severe Injury		Moderate Injury		PDO	
			Count	%	Count	%	Count	%
Road characteristic	1	Straight-leveled road	1675	56.7	50437	70.4	118417	72.4
	2	Straight-grade / hillcrest / bottom	485	16.4	12319	17.2	27794	17.0
	3	Curve-leveled / grade / hillcrest	788	26.7	8876	12.4	17221	10.5
	4	Not stated / unknown	4	0.1	56	0.1	180	0.1
Driver gender	1	Male	2006	68.0	38058	53.1	92305	56.4
	2	Female	946	32.0	33630	46.9	71307	43.6
Driver age	1	15–19 years	262	8.9	7105	9.9	16690	10.2
	2	19–69 years	2505	84.9	60725	84.7	138762	84.8
	3	≥70 years	185	6.3	3858	5.4	8160	5.0
Speed limit class	1	≤20 mph	5	0.2	415	0.6	1482	0.9
	2	20–30 mph* (30 mph included)	17	0.6	1064	1.5	3042	1.9
	3	30–40 mph	302	10.2	15354	21.4	35896	21.9
	4	40–50 mph	860	29.1	28690	40.0	62077	37.9
	5	50–60 mph	1495	50.6	20053	28.0	42518	26.0
	6	>60 mph	273	9.2	6112	8.5	18597	11.4
Crash type	1	Ran off-road	113	3.8	2524	3.5	5590	3.4
	2	Jackknife, overturn/rollover	124	4.2	1191	1.7	1276	0.8
	3	Pedestrian/pedal cyclist	186	6.3	484	0.7	53	0.0
	4	Animal or movable object	26	0.9	727	1.0	9287	5.7
	5	Parked vehicle or fixed object	664	22.5	9296	13.0	22011	13.5
	6	Rear-end collision	395	13.4	29527	41.2	67554	41.3
	7	Left-/right-turn crashes	340	11.5	9243	12.9	15346	9.4
	8	Head-on collision	416	14.1	1534	2.1	763	0.5
	9	Sideswipe or angle collision	599	20.3	15879	22.2	37251	2.8
	10	Other	89	3.0	1283	1.8	4481	2.7
Work zone area	0	No	2880	97.6	69872	97.5	159385	97.4
	1	Yes	72	2.4	1816	2.5	4227	2.6
Vehicle type	1	Passenger car/taxi	1331	45.1	40185	56.1	90650	55.4
	2	Pickup, light truck, sports utility, or van	1228	41.6	28305	39.5	65899	40.3
	3	Commercial bus, school bus, activity bus, other bus	13	0.4	255	0.4	593	0.4
	4	Single-unit truck, truck/trailer, truck/tractor, tractor doubles, semitrailer, farm equipment, or other heavy trucks	192	6.5	1847	2.6	5557	3.4
	5	Motor scooter, moped, pedal cycle, or motorcycle	179	6.1	876	1.2	185	0.1
	6	Other	9	0.3	220	0.3	728	0.4
Seasonal factors	1	Spring	623	21.1	18168	25.3	44375	27.1
	2	Summer	762	25.8	17669	24.6	37989	23.2
	3	Autumn	717	24.3	15408	21.5	33642	20.6
	4	Winter	850	28.8	20443	28.5	47606	29.1
Road terrain	1	Flat	748	25.3	13405	18.7	29761	18.2
	2	Rolling	1975	66.9	53124	74.1	120807	73.8
	3	Mountainous	229	7.8	5159	7.2	1304	8.0

Variable	Category	Description	Severe Injury		Moderate Injury		PDO	
			Count	%	Count	%	Count	%
Time of the day	1	12:00 AM–03:00 AM	193	6.5	2047	2.9	3992	2.4
	2	03:00 AM–06:00 AM	177	6.0	1880	2.6	4381	2.7
	3	06:00 AM–09:00 AM	412	14.0	11886	16.6	29002	17.7
	4	09:00 AM–12:00 PM	366	12.4	9488	13.2	22060	13.5
	5	12:00 PM–03:00 PM	445	15.1	13421	18.7	29605	18.1
	6	03:00 PM–06:00 PM	552	18.7	18351	25.6	41449	25.3
	7	06:00 PM–09:00 PM	499	16.9	10164	14.2	23404	14.3
	8	09:00 PM–12:00 PM	308	10.4	4451	6.2	9719	5.9
Day of the week	1	Sunday	412	14.0	6389	8.9	13217	8.1
	2	Monday	485	16.4	12518	17.5	29077	17.8
	3	Tuesday	410	13.9	12078	16.8	28917	17.7
	4	Wednesday	465	15.8	10533	14.7	24472	15.0
	5	Thursday	335	11.3	9869	13.8	22882	14.0
	6	Friday	449	15.2	12493	17.4	28711	17.5
	7	Saturday	396	13.4	7808	10.9	16336	10.0
Locality	1	Agricultural	1573	53.3	21185	29.6	47922	29.3
	2	Residential	597	20.2	13674	19.1	27615	16.9
	3	Commercial	760	25.7	35720	49.8	85696	52.4
	4	Institutional	8	0.3	625	0.9	1379	0.8
	5	Industrial	14	0.5	484	0.7	1000	0.6

In this study, the statistical models were setup to preserve the ordered nature of the severity class. Two ML models with different classification logics were applied, i.e., bagging and boosting techniques. The dataset was divided into training and test sets to evaluate the model's performance on unseen data. The training set was used to train the model, while the test set was used to assess its performance. The split ratio used in this study is 75:25 for the training and testing dataset, respectively. Since the dataset suffers from class imbalance, where one class has significantly fewer samples than the others, the Synthetic Minority Over-Sampling Technique for Nominal (SMOTE-N) data was applied to balance the classes. These methods help prevent the model from being biased towards the majority class and improve its ability to learn from the minority class.

3.2 Statistical Models

A class of logistic models known as ordered probability models, such as the OLM or OPM, proves useful for regression analysis when dealing with a modeled variable containing three or more categories, with crucial consideration of the order among these categories (Sasidharan & Menéndez, 2014). OLMs and OPMs are widely used to analyze and predict the relationship between ordinal dependent variables and a set of independent variables. OLMs and OPMs have been widely used to predict crash severity (Abdel-Aty, 2003; Ayuso & Santolino, 2007).

3.2.1 Ordered Logit Model (OLM)

The OLM was developed for weather-related crash severity analysis with severity levels defined as severe, moderate, and PDO.

Let Y_i be the ordinal response variable representing the crash severity of the i^{th} observation. The severity levels are coded as follows: severe ($Y_i = 2$), moderate ($Y_i = 1$), and PDO ($Y_i = 0$). The OLM can be represented as follows:

$$logit(P(Y_i \leq 1 \mid X_i)) = \alpha_1 + \beta_1 X_{i1} + \dots + \beta_p X_{ip}$$

$$logit(P(Y_i \leq 2 \mid X_i)) = \alpha_2 + \beta_1 X_{i2} + \dots + \beta_p X_{ip}$$

where

$P(Y_i \leq 1 \mid X_i)$ represents the probability that the crash severity of the *i-th* observation is less than or equal to moderate (level 1);

$P(Y_i \leq 2 \mid X_i)$ represents the probability that the crash severity of the *i-th* observation is less than or equal to severe (level 2);

X_i denotes the vector of explanatory variables for the *i-th* observation;

α_1 and α_2 are the intercepts specific to moderate and severe levels of crash severity, respectively; and

$\beta_1, \dots \beta_p$ are the coefficients associated with the explanatory variables $X_{i1}, \dots X_{ip}$, respectively.

Note that the reference category for the ordinal response variable is PDO (level 0), which does not have a corresponding logit equation. Its probability can be derived as 1 - $P(Y_i \leq 1 \mid X_i)$ since the sum of probabilities for all levels should equal 1.

3.2.2 Ordered Probit Model (OPM)

The OPM assumes that the relationship between a set of independent variables and the probability of each category of the dependent variable follows a standard normal distribution (Long, 1997). The OPM can be represented mathematically as follows:

$$P\left(Y \leq \frac{j}{X}\right) = \emptyset\left(\beta' X - \varphi_j\right)$$

where

$P\left(Y \leq \frac{j}{X}\right)$ represents the probability that the dependent variable *Y* takes a value less than or equal to *j*, given the value of the independent variables *X*;

$\emptyset$ represents the cumulative distribution function of the standard normal distribution;

β is a vector of coefficients;

X is a vector of independent variables; and

φ_j represents the threshold parameters delineating different categories of dependent variables.

3.3 Machine Learning (ML) Models

3.3.1 Random Forest Model (RF)

RF is an ensemble learning method that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Suppose an RF model consists of *N* decision trees (Biau, 2012). Each tree gives a classification, and the tree "votes" for that class. The forest chooses the classification having the most votes (over all the trees in the forest). Each tree is grown as follows:

- If the number of cases in the training set is *N*, then *N* cases are sampled at random—but with replacement from the original data. This sample will be the training set for growing the tree.
- If there are *M* input variables, a number *m* is specified such that at each node, *m* variables are selected at random out of the *M* and the best split on these *m* is used to split the node. The value of *m* is held constant during the forest growing.
- Each tree is grown to the largest extent possible and there is no pruning.

For a given test record, each tree in the forest gives a classification. The forest chooses the classification having the most votes (over all the trees in the forest), and in case of regression, it takes the average of outputs by different trees.

Mathematically, the prediction of an RF model for an input *x* can be written as

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} f_i(x)$$

where *fi*(*x*) is the prediction of the *i*-th decision tree.

The model would take as input features of a traffic incident (such as speed, weather condition, time of day, etc.), and output a severity class (severe injury, moderate injury, PDO). The model would be trained on a labeled dataset, and the aim would be to minimize the discrepancy between the predicted and actual labels. An RF's ability to combine multiple decision trees helps it avoid overfitting and generally results in a robust prediction performance.

3.3.2 Extreme Gradient Boosting Model (XGBoost)

XGBoost model is an ensemble of decision trees, where each subsequent tree tries to correct the errors made by the previous ones. It is an iterative process that aims to minimize a loss function (Ke et al., 2017; Charm et al., 2023). Mathematically, the prediction $\hat{y}$ of an XGBoost model with K trees for an input x can be written as follows:

$$\hat{y} = \sum_{k=1}^{K} f_k(x)$$

where $f_k(x)$ is the prediction of the k-th tree.

The objective function that XGBoost tries to minimize is represented as follows:

$$obj(\theta) = L(\theta) + \vartheta(\theta)$$

where θ represents the parameters of the model, $L(\theta)$ is the training loss function, and $\vartheta(\theta)$ is a regularization term that controls the complexity of the model.

The "gradient boosting" part of XGBoost comes from the fact that it trains each new tree to predict the negative gradient (or "residual") of the loss function with respect to the current predictions (Ke et al., 2017). This is why it is called "gradient boosting," as it uses gradient information to boost the performance of the ensemble. The model would learn to predict the severity of a crash based on input features such as weather condition, time of day, lighting condition, etc., by minimizing the discrepancy between its predictions and the actual labels, while also controlling the complexity of the model to prevent overfitting.

3.4 Deep Learning (DL) Models

3.4.1 Multi-layer Perceptron

MLP is a class of ANN composed of multiple layers of nodes (or "neurons") in a directed graph. Each layer is fully connected to the next one, meaning that each node in a given layer is connected to all nodes in the adjacent layers. Consider an input vector x of dimension d, which corresponds to the features of a given traffic incident. This includes factors such as lighting condition, weather condition, time of day, etc. The input is then passed through one or more hidden layers. Each node in a hidden layer computes a weighted sum of its inputs, adds a bias term, and applies an activation function. Mathematically, this can be expressed as follows.

$$h_i^{(l)} = f(\sum_j w_i^{(l)}.\, h_j^{(l-1)} + \, b_i^{(l)})$$

Here, $h_i^{(l)}$ is the output of the *i-th* node in the *l-th* layer, f is the activation function, $w_{i,j}^{(l)}$ is the weight connecting the *j-th* node in the *(l – 1)-th* layer to the *i-th* node in the *l-th* layer, and $b_i^{(l)}$ is the bias term for the *i-th* node in the *l-th* layer. The final hidden layer is fully connected to the output layer. In a classification context such as traffic crash severity prediction, the output layer would have one node for each class (severe injury, moderate injury, and PDO) and would use a SoftMax function to convert the outputs into probabilities summing to 1 (Sattar et al., 2023). The class with the highest probability is chosen as the prediction.

3.4.2 TabNet

The TabNet model is designed to learn and capture the complex relationships between weather-related features and the severity levels of crashes. It leverages various components to process and analyze the tabular data effectively. It consists of shared feature transformers, sparse attention mechanisms, sequential feature selection, and adversarial learning components. Shared feature transformers process the input features and extract meaningful representations. They learn patterns and relationships within the weather-related crash data, enabling the model to capture the important features that contribute to crash severity prediction.

Sparse attention mechanisms are incorporated in TabNet to selectively focus on relevant features while disregarding irrelevant ones. This attention mechanism identifies important patterns and interactions among the weather-related features. By attending to the most relevant features, TabNet can effectively learn the relationships between weather conditions and crash severity (Sattar et al., 2023). Sequential feature selection is a critical aspect of TabNet. It adaptively chooses the most informative features at each decision step, allowing the model to make accurate predictions (Jiang et al., 2022). This sequential selection process ensures that the relevant features are utilized for classifying the severity levels of weather-related crashes. To enhance model robustness and interpretability, TabNet incorporates adversarial learning. It includes an adversarial loss component that encourages the model to capture useful information from the dropped features while maintaining privacy and interpretability.

3.5 Model Evaluation and Comparison

Performance metrics are essential for assessing the effectiveness and reliability of predictive models in various contexts. The models are evaluated using metrics such as precision, recall, and F1 score. Precision measures the proportion of true positive predictions among all positive predictions, while recall measures the proportion of true positives among all actual positives. The F1 score is a balanced measure of precision and recall, and accuracy measures the proportion of correct predictions among all predictions. The model comparison was carried out by assessing the performance of different models to determine which model performs best. The best model is often the one that balances the trade-off between bias and variance and performs well on unseen data, demonstrating good generalization capabilities. The specific criteria for choosing the best model can depend on the context, including the relative costs of different types of prediction errors.

4. Results and Discussion

4.1. Imbalance Data Treatment

Wen et al. (2021) reviewed the current status on the use of ML in traffic crash prediction and identified data imbalance as a major issue. In this study, the SMOTE-N algorithm was employed in addressing class imbalance by generating synthetic samples for the minority class. The SMOTE-N algorithm is specifically designed for datasets with nominal predictor features. Figures 1 and 2 shows the visual representation of the dataset before and after data treatment.

Figure 1. Distribution of Crash Severity before Data Treatment

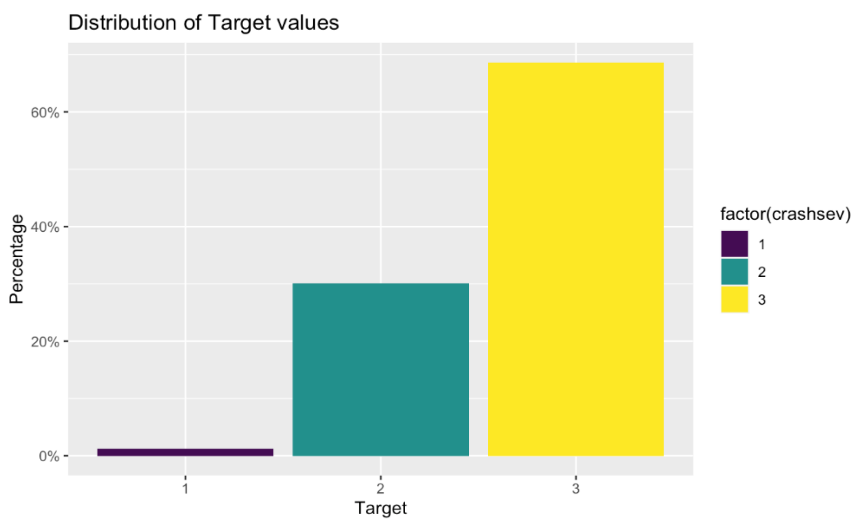
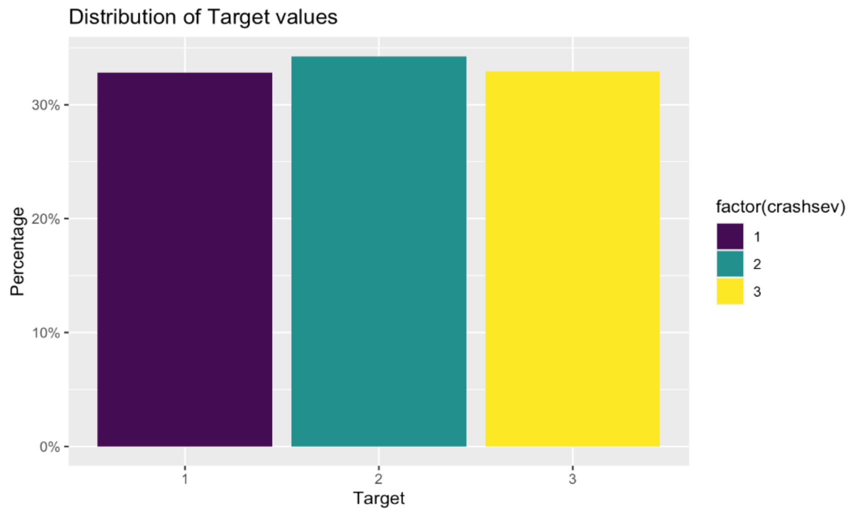


Figure 1 illustrates the data distribution wherein the minority class represents severe injury or class 1. Training an ML or DL model using such imbalanced datasets can lead to biased training samples. Consequently, the SMOTE-N technique was employed as a remedy for the data imbalance issue. Figure 2 visually depicts the distribution of the dataset subsequent to the application of the SMOTE-N data-treatment technique.

Figure 2. Distribution of Crash Severity after Data Treatment



4.2. Model Estimation

The main objective was to evaluate the performance of three different approaches (i.e., statistical, ML, and DL) to classify crash severity. For the ML and DL models, the SMOTE-N data treatment was used to resolve the data imbalance issue.

4.2.1. Statistical Models

4.2.1.1. Ordered Logit Model (OLM)

The confusion matrix from the OLM is as presented in Table 2. For severe crashes, the model did not correctly predict any instances. All severe crashes were misclassified as either moderate (215 instances) or PDO (525 instances). This indicates a major issue as the model is unable to correctly identify any of the severe crashes, which could be the most critical to predict accurately in a real-world setting. For moderate crashes, the model correctly predicted 1,101 instances. However, a significant number (16,776) were misclassified as PDO. This suggests that while the model has some ability to identify moderate crashes, it is largely confusing them with PDO crashes. The model performs best in predicting PDO crashes. It correctly predicted 40,195 PDO crashes and misclassified 751 as moderate. None of the PDO crashes were misclassified as severe. Hence, the model's performance in classifying crash severity is suboptimal. It never predicts severe crashes, which could be the most important to predict in practice. Additionally, it often confuses moderate and PDO crashes. The model may benefit from additional training, better feature selection, or a different modeling strategy.

Table 2. Confusion Matrix for the OLM

Test		Predicted		
		Severe Injury	Moderate Injury	PDO
Reference	Severe Injury	0	215	525
	Moderate Injury	0	1101	16776
	PDO	0	751	40195

4.2.1.2. Ordered Probit Model (OPM)

The confusion matrix from the OPM is as presented in Table 3. The model correctly predicted two moderate-injury crashes but also misclassified many crashes. Although the performance of both models is poor in this category, the OPM marginally outperformed the OLM. In addition, the model correctly predicted fewer moderate-injury crashes (1,022) and misclassified many as PDO. Therefore, the OLM slightly outperformed the OPM in this category. The model correctly predicted a slightly higher number of PDO crashes (40,410) and misclassified fewer as moderate-injury crashes (536). Thus, the OPM performs better in predicting PDO crashes. The performance of the OLM and OPM is relatively similar in this case, with each outperforming the other in different crash severity categories. However, both models struggle a great deal in correctly identifying moderate-injury crashes and distinguishing between moderate-injury and PDO crashes. These results suggest that further model refinement and exploration of other modeling techniques could be beneficial.

Table 3. Confusion Matrix for the OPM

Test		Predicted		
		Severe Injury	Moderate Injury	PDO
Reference	Severe Injury	2	236	502
	Moderate Injury	0	1022	16855
	PDO	0	536	40410

4.2.2. Machine Learning (ML) Models

4.2.2.1. Random Forest (RF)

Table 4 shows the confusion matrix from the RF model. Out of the actual severe-injury cases, the model correctly predicted 77 of them (true positives). However, it also misclassified 321 as moderate injury and 342 as PDO (false negatives). For the actual moderate injury cases, the model correctly classified 7,002 of them. On the other hand, it incorrectly classified 109 cases as severe injury and 10,766 cases as PDO. The model performed best in predicting the PDO crashes, correctly classifying 31,745 cases. However, it misclassified 154 cases as severe injury and 9,047 as moderate injury. While the model offers some value in identifying and categorizing crash severity, its current performance highlights several areas of potential concern in practical applications. These primarily revolve around the misclassification of cases, which could lead to inadequate responses, misallocation of resources, misguided policy decisions, and potential financial implications in an insurance context.

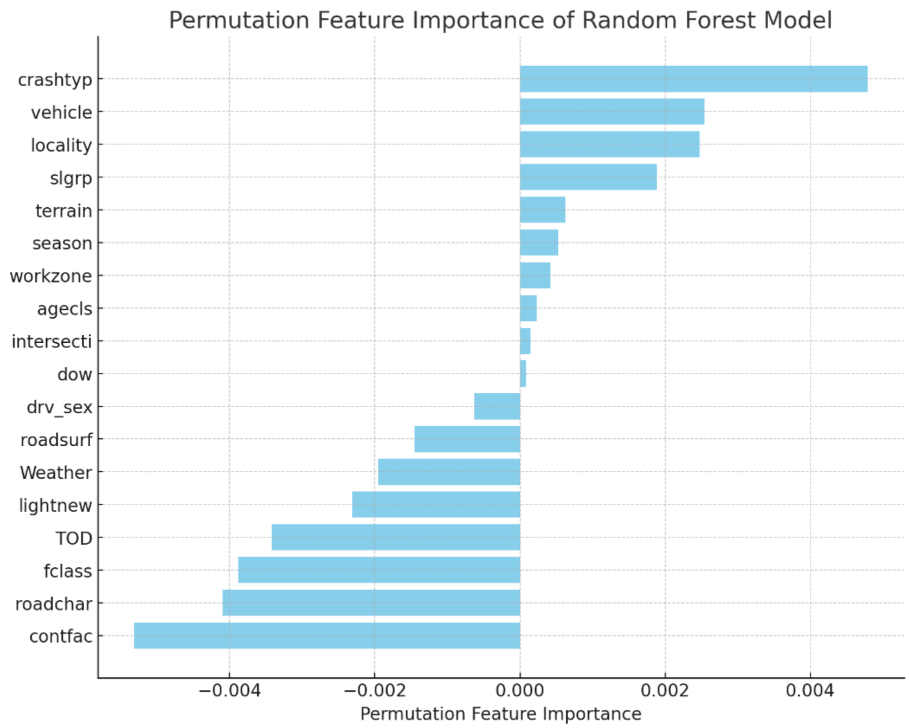
Table 4. Confusion Matrix for the RF

Test		**Predicted**		
		Severe Injury	**Moderate Injury**	**PDO**
Reference	**Severe Injury**	77	321	342
	Moderate Injury	109	7002	10766
	PDO	154	9047	31745

Figure 3 represents the permutation feature importance as determined by the RF model. The permutation feature importance is a technique for estimating the contributions of individual features to the predictive power of a model by observing the effect on model performance of randomly permuting the values of each feature, one at a time (Altmann et al., 2010). In this study, the RF model was trained on a subset of the dataset. A representative dataset was used for this process because this technique is computationally demanding. In the plot, each bar corresponds to a specific feature in the dataset, and the length of the bar corresponds to the importance of that feature. Positive values indicate that the performance of the model decreases when the feature is shuffled, suggesting that the model relies on the feature to make accurate predictions. Conversely, negative values indicate that the performance of the model improves when the feature is shuffled, suggesting that the model might be overfitting to noise in the feature.

The top three features according to this analysis are crash types, vehicle types, and locality, suggesting that these features are the most important for predicting the severity of a crash according to the trained model. However, it's important to note a few caveats. First, while permutation feature importance provides a useful way to rank the importance of features, it does not provide any information about the nature of the relationship between each feature and the target variable (Altmann et al., 2010). Second, it is possible that important features might appear unimportant if they are highly correlated with other features (Altmann et al., 2010). Finally, because this analysis was performed on a sample of the original dataset, the results might differ if the analysis were performed on the full dataset or a different sample.

Figure 3. Permutation Feature Importance Plot



4.2.2.1. Extreme Gradient Boosting (XGBoost)

Of the severe-injury cases, the model correctly predicted 139 of them (true positives). However, it misclassified 322 as moderate injury and 279 as PDO (false negatives). This model appears to be better at predicting severe-injury cases than the RF model, with more actual severe-injury cases correctly classified. For the moderate-injury cases, the model correctly classified 8,485 of them. However, it incorrectly classified 297 crashes as severe injury and 9,095 crashes as PDO.

Table 5. Confusion Matrix for the XGBoost

Test		Predicted		
		Severe Injury	**Moderate Injury**	**PDO**
Reference	**Severe Injury**	139	322	279
	Moderate Injury	297	8485	9095
	PDO	350	11999	28597

4.3. Model Performance Comparison and Analysis of Predictive Power

Table 6 presents the performance metrics of all the developed models. Evaluating model performance is a critical part of the ML process (Ke et al., 2017). By quantifying how well a model performs on a given dataset, one can understand its strengths and weaknesses, guide the selection of models, tune hyperparameters, and assess the effectiveness of feature selection or engineering (Ke et al., 2017). Without this step, one would not know if our model is effective, if it generalizes well to unseen data, or how it might be improved. When dealing with imbalanced datasets, the choice of evaluation metric becomes particularly important. In this study, the F1 score, alongside precision and recall, was used.

Table 6. Model Performance Metrics

Method	Model	Level of Severity	Precision %	Recall %	F1 Score %
Statistical Model	OLM	Severe Injury	Undefined	0.00	Undefined
		Moderate Injury	53.27	6.16	11.04
		PDO	69.91	98.17	81.66
	OPM	Severe Injury	100	0.27	0.54
		Moderate Injury	56.97	5.72	10.39
		PDO	69.95	98.69	81.87
ML	RF	Severe Injury	22.65	10.41	14.26
		Moderate Injury	42.77	85.53	57.03
		PDO	95.72	77.53	85.67
	XGBoost	Severe Injury	17.68	18.78	18.22
		Moderate Injury	40.78	47.46	43.87
		PDO	75.71	69.84	72.47
DL	MLP	Severe Injury	11.17	36.91	17.15
		Moderate Injury	5.08	60.44	9.37
		PDO	98.80	70.02	81.75
	TabNet	Severe Injury	8.45	45.73	28.78
		Moderate Injury	3.73	72.19	14.54
		PDO	79.62	82.72	92.10

4.3.1. Severe-injury classification

The most effective model for predicting severe-injury crashes is reportedly the TabNet model, which presumably is a form of DL model. This model has a precision of 8.45, which indicates that out of all instances predicted as severe-injury crashes, only 8.45% were severe-injury crashes. The recall is 45.73, meaning that the model correctly identified 45.73% of the actual severe-injury crashes. The F1 score, which is a harmonic mean of precision and recall, and is often used as a single metric to compare models, is 28.78. This relatively low score indicates that the model struggles with the severe-injury classification, which could be due to several reasons such as class imbalance or the complexity of the problem. However, when it comes to predicting crash severity, other studies have found XGBoost to perform better (Guo et al., 2021; Jamal et al., 2021). For instance, a comparative study by Jamal et al. (2021) found that an XGBoost model outperformed other ML models in predicting crash severity. In another study, Pensantez-Narvaez et al. (2019) compared the XGBoost model and logistic regression in crash severity analysis and found XGBoost to perform better. They attributed its success to its ability to handle non-linear relationships and interactions between features, which are common in crash data. While other studies might have found XGBoost to be a better model, the choice of model should be guided by the specific characteristics of the dataset and the problem at hand, and no single model is likely to be the best choice for all tasks (Li et al., 2023).

4.3.2. Moderate-injury classification

The model with the best performance for predicting moderate-injury crashes is RF. However, it has a precision of 42.77, which means that 42.77% of the instances it predicted as moderate-injury crashes were indeed moderate-injury crashes. It has a recall of 85.53, signifying that it correctly identified 85.53% of all actual moderate-injury crashes. The F1 score is 57.03, which is considerably higher than the severe-injury model, indicating a better balance of precision and recall. In previous research on crash severity prediction (Theofilatos et al., 2019), achieving a balance between precision and recall is a common challenge, as the relative importance of false positives and false negatives can vary depending on the specific application and costs associated with prediction errors. The RF model's high recall indicates its ability to capture a substantial proportion of actual moderate injuries, a crucial consideration in traffic safety applications.

4.3.3. Property Damage Only (PDO) Classification

The model with the highest performance in predicting PDO crashes is TabNet. Its precision is 79.62, implying that 79.62% of instances it predicted as PDO were PDO. The model has a recall of 82.72, implying it correctly identified 82.72% of actual PDO cases. The F1 score is notably high at 92.1, indicating a superior balance of precision and recall for this class.

Hence, the model performances vary significantly across the severity levels, with the PDO predictions faring the best and the severe-injury predictions performing the worst. This discrepancy in performance may be attributed to various factors, including but not limited to, class imbalance, differing complexities in predicting each class, or nuances in the data distribution for each class.

Figure 4. Precision Comparison for Statistical, ML, and ML Models

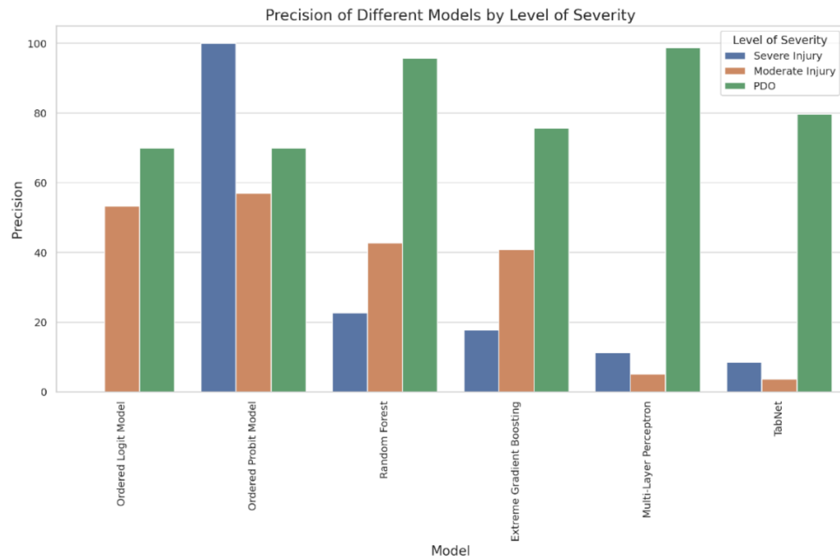


Figure 5. Recall Comparison for Statistical, ML, and DL Models

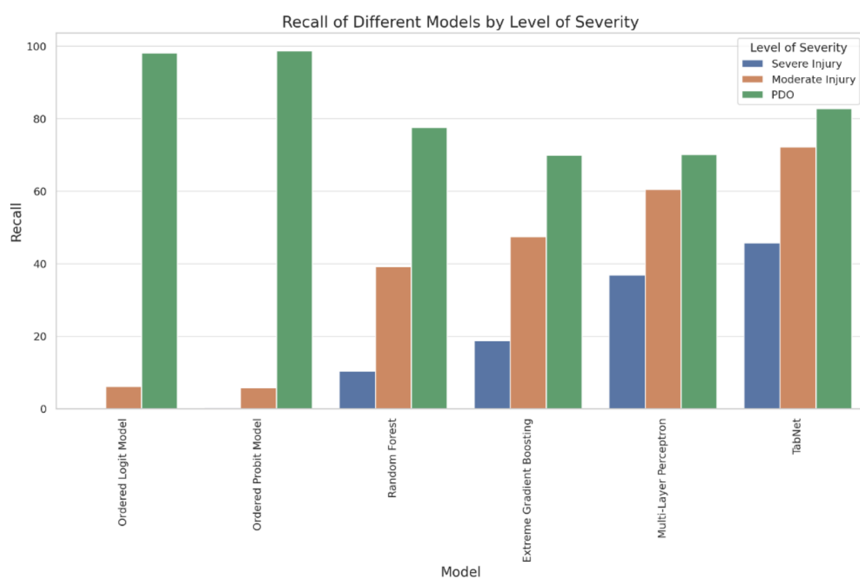


Figure 6. F1 Comparison for Statistical, ML, and DL Models

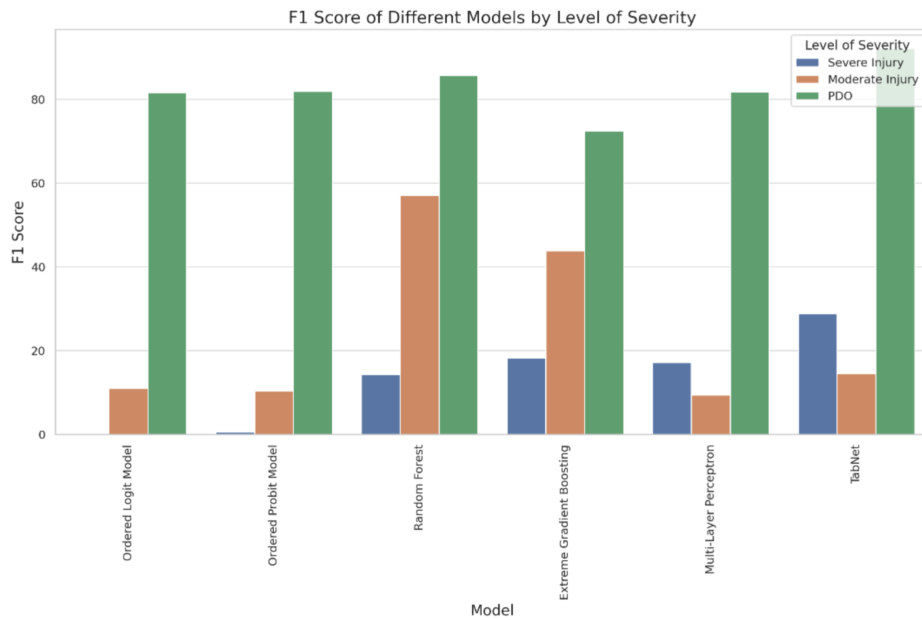


Figure 6 shows the graphical comparison of the performance metrics of different models. The statistical modes show high precision for severe-injury crashes but very low recall, suggesting they are confident but not comprehensive in identifying severe-injury crashes. The F1 score, which balances precision and recall, is also low for severe-injury crashes. For moderate-injury crashes, precision is relatively high, but recall is low, leading to a low F1 score. However, these models perform better for PDO crashes, with a relatively high F1 score. The ML models, on the other hand, provide a more balanced performance across all severity levels. For severe-injury crashes, both precision and recall are lower than for other categories, resulting in a lower F1 score. For moderate injury and PDO crashes, the performance is better with higher F1 scores. The DL models, however, have lower precision for moderate- and severe-injury crashes, suggesting they may incorrectly classify incidents as these categories. However, they have high recall, particularly for PDO and moderate-injury crashes, indicating that they are good at identifying these categories when present. The F1 scores are relatively low for severe- and moderate-injury crashes, but higher for PDO crashes.

While the statistical models are very confident in their predictions of severe-injury crashes (high precision), they miss a lot of actual severe-injury crashes (low recall). This might be acceptable if the goal is to avoid false alarms at the cost of missing some real cases. However, in many applications, such as emergency response or healthcare, missing severe-injury crashes could have serious consequences, making these models less effective for real-world application. For predicting PDO crashes, statistical models perform very well with a high recall and good precision. The ML models offer a more balanced performance, making fewer mistakes in both directions (moderate precision and recall). They may be useful in situations where both false positives and false negatives carry significant costs. For example, if resources are being allocated based on these predictions, it would be important to both catch as many real cases as possible (high recall) and to avoid wasting resources on false alarms (high precision). The DL models, however, are very sensitive, catching a large proportion of actual incidents (high recall), but they also produce a significant number of false alarms (low precision). This could be useful in a situation where the cost of a false negative is much higher than the cost of a false positive. For example, in a preliminary screening tool, it might be acceptable to have more false alarms to ensure that all or most real cases are identified for further investigation.

5. Conclusion

5.1 Summary of Findings

An extensive evaluation of various statistical, ML, and DL models to predict crash severity was conducted in this study. The analysis began with data preprocessing using SMOTE-N to address imbalances in the crash dataset. The findings revealed that each model category presented unique strengths and weaknesses. Statistical models, including OLM and OPM, demonstrated high precision but low recall in predicting severe-injury crashes. These models excelled in their confidence but lacked comprehensive identification of severe incidents. They were most effective in predicting PDO crashes.

ML models, exemplified by RF and XGBoost, delivered a more balanced performance. They provided moderate precision and recall across different severity levels, making them suitable for situations where the costs of false positives and false negatives are significant. DL models, MLP, and TabNet offered high recall rates but lower precision. These models were particularly effective in identifying severe- and moderate-injury crashes, despite a higher rate of false positives. TabNet stood out as the best performer for predicting both severe-injury and PDO crashes. It can be deduced from this study that the choice of model heavily depends on the specific application and the trade-off between false positives and false negatives. While statistical models provide high precision, ML models offer balanced performance, and DL models yield high recall. Therefore, to optimally predict crash severity, practitioners should carefully consider the context and requirements of their specific application.

The models developed in this study hold substantial promise for enhancing traffic safety in real-world applications. Specifically, the RF model exhibits notable performance in predicting moderate-injury crashes, offering a valuable tool for traffic safety authorities and organizations. Its balanced precision and recall make it suitable for identifying crashes with a moderate level of injury severity, allowing for more targeted allocation of resources, such as emergency response teams and medical personnel, to the crash scenes. This can lead to quicker and more effective responses, potentially reducing the severity of injuries and saving lives. Furthermore, the TabNet model's exceptional precision and recall in predicting PDO crashes make it a valuable asset for traffic management and insurance agencies. By accurately identifying PDO crashes, these organizations can streamline their claims processes, minimize fraud, and allocate resources more efficiently.

Moreover, the deployment of these models can significantly contribute to proactive traffic safety measures. By leveraging the insights gained from predictive modeling, traffic safety authorities can implement targeted interventions and educational campaigns in areas prone to severe crashes. Additionally, these models can inform the design and implementation of infrastructure improvements, such as road modifications and signage enhancements, to mitigate crash risks effectively. As traffic safety continues to be a paramount concern worldwide, the integration of predictive models such as RF and TabNet into decision-making processes can lead to more effective and data-driven approaches to reducing injuries and fatalities on our roadways. Ultimately, the utilization of these models empowers stakeholders with the tools they need to make more informed, proactive, and lifesaving decisions in the field of traffic safety.

5.2 Limitations and Future Scope of Work

The study provides valuable insights into the performance of various models for predicting crash severity. The following are limitations and should be further explored:

- Imbalanced data were dealt with using the SMOTE-N technique. Although this method helps balance the classes, it also created synthetic data, which might not represent real-world scenarios accurately. There is a need to further explore the sensitivity of different data-treatment methods to crash data.

- There is a need for further refinement of the developed models, especially in correctly identifying moderate-injury crashes and distinguishing between moderate injury and PDO crashes.

- ML techniques have shown promise in improving prediction accuracy, while statistical models provide interpretability. Future research should focus on addressing data limitations, integrating real-time data sources, and exploring emerging technologies such as sensor data fusion and DL algorithms to enhance crash severity prediction and improve road safety outcomes.

Abbreviations and Acronyms

ANN	Artificial Neural Network
AUC-ROC	Area Under Curve- Receiver Operating Characteristic
CNN	Convolutional Neural Network
DL	Deep Learning
HSIS	Highway Safety Information System
ML	Machine Learning
MLP	Multi-Layer Perceptron
OLM	Ordered Logit Model
OPM	Ordered Probit Model
PDO	Property Damage Only
ResNet	Residual Neural Network
RF	Random Forest
RNN	Recurrent Neural Network
XGBoost	Extreme Gradient Boosting

References

1. Gopalakrishnan, S. (2012). A public health perspective of road traffic accidents. *Journal of Family Medicine and Primary Care*, *1*(2), 144.

2. Centers for Disease Control and Prevention (CDC). (2023, January 20). *Transportation safety*. https://www.cdc.gov/transportationsafety/index.html

3. Druta, C., Kassing, A., Gibbons, R., & Alden, V. A. (2020). Assessing driver behavior using shrp2 adverse weather data. *Journal of Safety Research*, *73*, 283–295

4. Greaves, H. E., Boelman, N. T., Brinkman, T. J., Liston, G. E., Prugh, L. R., & Reinking, A. K. (2023). Simulating future climate change impacts on snow- and ice-related driving hazards in Arctic-boreal regions. *Environmental Research Letters*, *18*(2), 025006.

5. Federal Highway Administration (FHWA). (2020). *How do weather events impact roads?* https://ops.fhwa.dot.gov/weather/q1_roadimpact.htm

6. Penmetsa, P., & Pulugurtha, S. S. (2017). Risk drivers pose to themselves and other drivers by violating traffic rules. *Traffic Injury Prevention Journal*, *18*(1), 63–69.

7. Penmetsa, P., & Pulugurtha, S. S. (2019). Risk factors attributed to fatal fixed object crashes on non-interstate roads. *Journal of Transportation Safety & Security*, *11*(1), 102–116.

8. Penmetsa, P., Pulugurtha, S. S., & Duddu, V. R. (2018a). Factors associated with crashes due to overcorrection or oversteering of vehicles. *International Association of Traffic and Safety Sciences (IATSS) Research Journal*, *42*(1), 24–29.

9. Tagar, S., & Pulugurtha, S. S. (2021). Predictor variables influencing merging speed change lane crash risk by interchange type in urban areas. *Transportation Research Interdisciplinary Perspectives Journal*, *10*, 100375.

10. Chen, F., Song, M., & Ma, X. (2019). Investigation on the injury severity of drivers in rear-end collisions between cars using a random parameters bivariate ordered probit model. *International Journal of Environmental Research and Public Health*, *16*(14), 2632.

11. Fanyu, M., Sze, N. N., Cancan, S., Tiantian, C., & Yiping, Z. (2021). Temporal instability of truck volume composition on non-truck-involved crash severity using uncorrelated and correlated grouped random parameters binary logit models with space-time variations. *Analytic Methods in Accident Research*, *31*, 100168.

12. Hou, Q., Huo, X., Leng, J., & Mannering, F. (2022). A note on out-of-sample prediction, marginal effects computations, and temporal testing with random parameters crash-injury severity models. *Analytic Methods in Accident Research*, *33*, 100191.

13. Mao, X., Yuan, C., Gan, J., & Zhang, S. (2019). Risk factors affecting traffic accidents at urban weaving sections: Evidence from China. *International Journal of Environmental Research and Public Health*, *16*(9), 1542.

14. Abdulhafedh, A. (2022). Comparison between common statistical modeling techniques used in research, including: Discriminant analysis vs. logistic regression, ridge regression vs. LASSO, and decision tree vs. random forest. *Open Access Library Journal*, *9*(2), 1–19.

15. Mokhtarimousavi, S., Anderson, J. C., Azizinamini, A., & Hadi, M. (2019). Improved support vector machine models for work zone crash injury severity prediction and analysis. *Transportation Research Record*, *2673*(11), 680–692.

16. Theofilatos, A., Chen, C., & Antoniou, C. (2019). Comparing machine learning and deep learning methods for real-time crash prediction. *Transportation Research Record*, *2673*(8), 169–178.

17. Hadjidimitriou, N. S., Lippi, M., Dell'Amico, M., & Skiera, A. (2019). Machine learning for severity classification of accidents involving powered two wheelers. *IEEE Transactions on Intelligent Transportation Systems*, *21*(10), 4308–4317.

18. Al-Mistarehi, B. W., Alomari, A. H., Imam, R., & Mashaqba, M. (2022). Using machine learning models to forecast severity level of traffic crashes by R Studio and ArcGIS. *Frontiers in Built Environment*, *8*, 860805.

19. Zhao, S., Wang, K., Liu, C., & Jackson, E. (2019). Investigating the effects of monthly weather variations on Connecticut freeway crashes from 2011 to 2015. *Journal of Safety Research*, *71*, 153–162.

20. Wei, Z., Zhang, Y., & Das, S. (2023). Applying explainable machine learning techniques in daily crash occurrence and severity modeling for rural interstates. *Transportation Research Record*, 03611981221134629.

21. Zhang, X., Wen, H., Yamamoto, T., & Zeng, Q. (2021). Investigating hazardous factors affecting freeway crash injury severity incorporating real-time weather data: Using a Bayesian multinomial logit model with conditional autoregressive priors. *Journal of Safety Research*, *76*, 248–255.

22. Hammad, H. M., Ashraf, M., Abbas, F., Bakhat, H. F., Qaisrani, S. A., Mubeen, M., ... & Awais, M. (2019). Environmental factors affecting the frequency of road traffic accidents: a case study of sub-urban area of Pakistan. *Environmental Science and Pollution Research*, *26*, 11674–11685.

23. Das, S., Dutta, A., & Sun, X. (2020). Patterns of rainy weather crashes: Applying rules mining. *Journal of Transportation Safety & Security*, *12*(9), 1083–1105.

24. Robbins, C. J., & Fotios, S. (2020). Motorcycle safety after-dark: the factors associated with greater risk of road-traffic collisions. *Accident Analysis & Prevention*, *146*, 105731.

25. Xing, F., Huang, H., Zhan, Z., Zhai, X., Ou, C., Sze, N. N., & Hon, K. K. (2019). Hourly associations between weather factors and traffic crashes: Non-linear and lag effects. *Analytic Methods in Accident Research*, *24*, 100109.

26. Abohassan, A., El-Basyouny, K., & Kwon, T. J. (2021). Exploring the associations between winter maintenance operations, weather variables, surface condition, and road safety: A path analysis approach. *Accident Analysis & Prevention*, *163*, 106448.

27. Yang, Y., Wang, K., Yuan, Z., & Liu, D. (2022). Predicting freeway traffic crash severity using XGBoost-Bayesian network model with consideration of features interaction. *Journal of Advanced Transportation*, *2022*.

28. Druta, C., Kassing, A., Gibbons, R., & Alden, V. A. (2020). Assessing driver behavior using shrp2 adverse weather data. *Journal of Safety Research*, *73*, 283–295.

29. Duddu, V. R., Pulugurtha, S. S., Mane, A. S., & Godfrey, C. (2020). Back-propagation neural network model to predict visibility at a road link-level. *Transportation Research Interdisciplinary Perspectives*, *8*, 100250.

30. Mathew, S., & Pulugurtha, S. S. (2022). *Effect of weather events on travel time reliability and crash occurrence. [Report]* (No. 22-41, CA-MTI-2035). San Jose State University. College of Business. Mineta Transportation Institute.

31. Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences*, *113*(10), 2636–2641.

32. Sattar, K., Chikh Oughali, F., Assi, K., Ratrout, N., Jamal, A., & Masiur Rahman, S. (2023). Transparent deep machine learning framework for predicting traffic crash severity. *Neural Computing and Applications*, *35*(2), 1535–1547.

33. Duddu, V. R., Kukkapalli, V. M., & Pulugurtha, S. S. (2019). Crash risk factors associated with injury severity of teen drivers. *IATSS Research*, *43*(1), 37–43.

34. Islam, M., & Mannering, F. (2020). A temporal analysis of driver-injury severities in crashes involving aggressive and non-aggressive driving. *Analytic Methods in Accident Research*, *27*, 100128.

35. Shi, X., Wong, Y. D., Li, M. Z. F., Palanisamy, C., & Chai, C. (2019). A feature learning approach based on XGBoost for driving assessment and risk prediction. *Accident Analysis & Prevention*, *129*, 170–179.

36. Yuan, J., Abdel-Aty, M., Gong, Y., & Cai, Q. (2019). Real-time crash risk prediction using long short-term memory recurrent neural network. *Transportation Research Record*, *2673*(4), 314–326.

37. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, *30*.

38. 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. *International Journal of Injury Control and Safety Promotion*, *28*(4), 408–427.

39. Penmetsa, P., & Pulugurtha, S. S. (2018b). Modeling crash injury severity by road feature to improve safety. *Traffic Injury Prevention*, *19*(1), 102–109.

40. Yan, X., He, J., Zhang, C., Liu, Z., Wang, C., & Qiao, B. (2021). Spatiotemporal instability analysis considering unobserved heterogeneity of crash-injury severities in adverse weather. *Analytic Methods in Accident Research*, *32*, 100182.

41. Ayuso, M., & Santolino, M. (2007). Predicting automobile claims bodily injury severity with sequential ordered logit models. *Insurance: Mathematics and Economics*, *41*(1), 71–83.

42. Abdel-Aty, M. (2003). Analysis of driver injury severity levels at multiple locations using ordered probit models. *Journal of Safety Research*, *34*(5), 597–603.

43. Gray, R. C., Quddus, M. A., & Evans, A. (2008). Injury severity analysis of accidents involving young male drivers in Great Britain. *Journal of Safety Research*, *39*(5), 483–495.

44. Garrido, R., Bastos, A., de Almeida, A., & Elvas, J. P. (2014). Prediction of road accident severity using the ordered probit model. *Transportation Research Procedia*, *3*, 214–223.

45. Penmetsa, P., Pulugurtha, S. S., & Duddu, V. R. (2017). Examining injury severity of not at-fault drivers in two-vehicle crashes. *Journal of Transportation Research Board, National Research Council, Transportation Research Record Series 2659*, 164–173.

46. Xie, Y., Zhang, Y., & Liang, F. (2009). Crash injury severity analysis using Bayesian ordered probit models. *Journal of Transportation Engineering*, *135*(1), 18–25.

47. Behnood, A., & Mannering, F. L. (2015). The temporal stability of factors affecting driver-injury severities in single-vehicle crashes: Some empirical evidence. *Analytic Methods in Accident Research*, *8*, 7–32.

48. Behnood, A., & Mannering, F. L. (2016). An empirical assessment of the effects of economic recessions on pedestrian-injury crashes using mixed and latent-class models. *Analytic Methods in Accident Research*, *12*, 1–17.

49. Behnood, A., & Mannering, F. (2017). The effect of passengers on driver-injury severities in single-vehicle crashes: A random parameters heterogeneity-in-means approach. *Analytic Methods in Accident Research*, *14*, 41–53.

50. Chen, C., Zhang, G., Huang, H., Wang, J., & Tarefder, R. A. (2016). Examining driver injury severity outcomes in rural non-interstate roadway crashes using a hierarchical ordered logit model. *Accident Analysis & Prevention*, *96*, 79–87.

51. Fountas, G., & Anastasopoulos, P. C. (2018). Analysis of accident injury-severity outcomes: The zero-inflated hierarchical ordered probit model with correlated disturbances. *Analytic Methods in Accident Research*, *20*, 30–45.

52. Duddu, V. R., Penmetsa, P., & Pulugurtha, S. S. (2018). Modeling and comparing injury severity of at-fault and not at-fault drivers in crashes. *Accident Analysis & Prevention*, *120*, 55–63.

53. Duvvuri, S., Pulugurtha, S. S., & Mathew, S. (2022). Modeling injury severity of crashes involving trucks: Capturing and exploring risk factors associated with land use and demographic in addition to crash, driver, and on-network characteristics. *IATSS Research*, *46*(4), 602–613.

54. Zheng, M., Li, T., Zhu, R., Chen, J., Ma, Z., Tang, M., Cui Z. & Wang, Z. (2019). Traffic accident's severity prediction: A deep-learning approach-based CNN network. *IEEE Access*, *7*, 39897–39910.

55. Wang, L., Abdel-Aty, M., Lee, J., & Shi, Q. (2019). Analysis of real-time crash risk for expressway ramps using traffic, geometric, trip generation, and socio-demographic predictors. *Accident Analysis & Prevention*, *122*, 378–384.

56. Santos, K., Dias, J. P., & Amado, C. (2022). A literature review of machine learning algorithms for crash injury severity prediction. *Journal of Safety Research*, *80*, 254–269.

57. Tang, J., Zheng, L., Han, C., Yin, W., Zhang, Y., Zou, Y., & Huang, H. (2020). Statistical and machine-learning methods for clearance time prediction of road incidents: A methodology review. *Analytic Methods in Accident Research*, *27*, 100123.

58. Khan, M. N., & Ahmed, M. M. (2022). A novel deep learning approach to predict crash severity in adverse weather on rural mountainous freeway. *Journal of Transportation Safety & Security*, 1–31.

59. Ma, Z., Mei, G., & Cuomo, S. (2021). An analytic framework using deep learning for prediction of traffic accident injury severity based on contributing factors. *Accident Analysis & Prevention*, *160*, 106322.

60. Rahim, M. A., & Hassan, H. M. (2021). A deep learning-based traffic crash severity prediction framework. *Accident Analysis & Prevention*, *154*, 106090.

61. Das, S., Dutta, A., Dixon, K., Minjares-Kyle, L., & Gillette, G. (2018). Using deep learning in severity analysis of at-fault motorcycle rider crashes. *Transportation Research Record*, *2672*(34), 122–134.

62. Alkheder, S., Taamneh, M., & Taamneh, S. (2017). Severity prediction of traffic accident using an artificial neural network. *Journal of Forecasting*, *36*(1), 100–108.

63. 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. *Transportation Research Record*, *1746*(1), 6–13.

64. Tan, Mingxing, and Quoc Le. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*, 6105–6114.

65. Washington, S., & Haque, M. D. (2013). On the commonly accepted assumptions regarding observed motor vehicle crash counts at transport system locations. In *Transportation Research Board (TRB) 92nd Annual Meeting Compendium of Papers*, 1-19. Transportation Research Board (TRB), National Academy of Sciences.

66. Sasidharan, L., & Menéndez, M. (2014). Partial proportional odds model—An alternate choice for analyzing pedestrian crash injury severities. *Accident Analysis & Prevention*, *72*, 330–340.

67. Long, J. S. (1997). Regression models for categorical and limited dependent variables (Vol. 7). *Advanced Quantitative Techniques in the Social Sciences*, *219*.

68. Biau, G. (2012). Analysis of a random forests model. *The Journal of Machine Learning Research*, *13*, 1063–1095.

69. Charm, T., Wang, H., Zuniga-Garcia, N., Ahmed, M., & Kockelman, K. M. (2023). Predicting crash occurrence at intersections in Texas: an opportunity for machine learning. *Transportation Planning and Technology*, 1–21.

70. Jiang, T., Hu, W., Liu, Y., & Xiao, L. (2022, November). Capturing and Exploring Unobserved Heterogeneity in Traffic Crash Injuries in China: A Deep Learning Approach. In *2022 IEEE 7th International Conference on Intelligent Transportation Engineering (ICITE)*, 331–336. IEEE.

71. 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. *Transport Reviews*, *41*(6), 855–879.

72. Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: a corrected feature importance measure. *Bioinformatics*, *26*(10), 1340–1347.

73. Guo, M., Yuan, Z., Janson, B., Peng, Y., Yang, Y., & Wang, W. (2021). Older pedestrian traffic crashes severity analysis based on an emerging machine learning XGBoost. *Sustainability*, *13*(2), 926.

74. Pesantez-Narvaez, J., Guillen, M., & Alcañiz, M. (2019). Predicting motor insurance claims using telematics data—XGBoost versus logistic regression. *Risks*, *7*(2), 70.

75. Li, H., Fowlkes, C., Yang, H., Dabeer, O., Tu, Z., & Soatto, S. (2023). Guided recommendation for model fine-tuning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3633–3642.

About the Authors

Abimbola Ogungbire, M.S.

Mr. Ogungbire is pursuing his Ph.D. in the Infrastructure and Environmental Systems (INES) program at the University of North Carolina at Charlotte. His areas of interest are traffic data analytics and visualization.

Panick Kalambay, Ph.D.

Dr. Kalambay recently graduated with his doctoral degree in Civil Engineering at the University of North Carolina at Charlotte. His areas of interest are traffic operations & safety, and transportation planning.

Hardik Gajera, Ph.D.

Dr. Gajera recently graduated with his doctoral degree in Civil Engineering at the University of North Carolina at Charlotte. His areas of interests include traffic operations & safety, traffic simulations, and connected & automated vehicles.

Srinivas S. Pulugurtha, Ph.D., P.E., F.ASCE

Dr. Pulugurtha is currently working as a Professor & Research Director of the Department of Civil and Environmental Engineering at the University of North Carolina at Charlotte.

MTI FOUNDER

Hon. Norman Y. Mineta

MTI BOARD OF TRUSTEES

Directors

MTI MINETA TRANSPORTATION INSTITUTE