Enhance Road Safety, Focusing on Pedestrian: Using Taipei Traffic Accident Data in 2020-2022

Ling-Yun, Huang
School of Science & Technology
Department of Computer Science
City, University of London
London, United Kindom
Ling-Yun.Huang@city.ac.uk

***Abstract*—This study analyses the Taipei traffic accident data from 2020 to 2022, specifically focusing on pedestrian-related accidents. Analytical methods, including chi-squared tests, logistic regression, heatmap, and the Random Forest model, uncover the key factors that influence pedestrian safety. Factors correlated with fatality accidents include timing, location and vehicle type, indicating the higher risks in bus or complex accidents and at midnight danger increased. Seasonal pattern shows more accidents in winter evenings compared to other periods, contributed by combination with darker light conditions, adverse weather, and rush hours. In analysing pedestrian injury severity, age is the most important factor with older people having a higher possibility of fatal condition. Other factors such as the driver's age, cause of the accident, driver's vehicle, and time of the accident also play a role in injury severity. The findings of this study offer valuable insights to enhance pedestrian safety.**

Keywords—pedestrian, safety, traffic accident

I. Introduction

Based on a CNN report on December 6, 2022, Taiwan's traffic issues have generated significant attention, and these issues highly affect the tourism industry and cause challenges to road safety [1]. The United States government has warned its citizens to take notice of road safety while travelling on the island in their travelling guidelines [2]. This doesn't just influence the flow of tourists; it also appears a major concern for the Taiwanese people.

According to the Taiwanese government's record in 2022, there were 17,742 pedestrians involved in traffic accidents, including 16,945 injuries and 432 fatalities [3]. The pedestrian injury rate is 74.7 per 100,000 population and the pedestrian fatality rate is 1.9 per 100,000 population. In comparison to the study by E. Charters et al in 2017, other countries like Sweden, France, and Australia show better rates in similar issues [4].

This report used the data based on Taipei City. With its well-structured public transportation systems and many people working and living in this city, Taipei has a higher pedestrian density than other cities in Taiwan. Analysing Taipei's traffic accident data could provide valuable insights for enhancing road safety, particularly for pedestrians. This research aims to find possible risk factors for fatal accidents involving pedestrians and explore why certain time periods may cause more accidents or which factors may play a more significant role in pedestrian injury severity.

II. Analytical Questions

This study used datasets that provided information at the individual level. Also, it included all traffic accidents that involved injury or death in Taipei from 2020 to 2022. It allows us to analyse data from both accidents and individual perspectives. To assess the overall impact, first, we would want to know the fatal or injury rate with different types of vehicles involved.

Following the aim of this study, we would shift our focus to pedestrian data. The analytical questions are as follows:

- Are the fatal and injury rates in pedestrians higher than other road users?
- In pedestrian-related accidents, are there seasonal, monthly, or hourly patterns, weather or road conditions, location, and human behaviour associated with fatal incidents compared to non-fatal ones?
- Are there any periods of time that might have a higher number of pedestrian-related accidents, and what factors might cause this situation?
- Which factors such as age, gender, accident conditions, and driver's information may play a more significant role in pedestrian injury severity? And how?

By answering all these analytical questions, we could have a chance to look deeper into pedestrian accidents. And might gain useful insight to reduce or prevent pedestrian-related accidents or increase its survival rate.

III. Data (Materials)

The data in this research originates from the records supported by the Taipei City Police Department and is available through Taiwan's open data platform. This report used the A1 and A2 types of traffic accidents in Taipei from 2020 to 2022 [5]. There are three categories related to traffic accidents in Taiwan, the A1 category involves traffic accidents resulting in fatalities within 24 hours, the A2 category includes incidents leading to fatalities within 30 days or injuries, and the A3 category includes incidents that only involve vehicle damage and property loss. The dataset only focuses on the first two categories.

These datasets were documented at the individual level, including age, gender, vehicle type, injury severity, and human behaviour (e.g., driver's qualification, and alcohol use). And included accident details such as dates and times, location, weather, light, road conditions, and the main cause of the accident etc. However, some variables may require transformation for later analysis since there are too many groups in one column or too few numbers in some groups.

Across all three years of data, that includes 166,818 individuals with 75,229 traffic accidents leading to 263 fatalities and 99,751 injuries. Among all accidents, 6,282 incidents involved pedestrians with 101 fatalities. All the details above give us enough features and numbers to answer our analytical questions.

IV. Analysis

A. Data Preparation

The datasets from the platform are provided annually and with their variables' names in Mandarin. All features have been translated into English before import. Since this study focuses on human-related factors, rows referred to as property or unknown will be excluded for the later analysis.

B. Variable Transformations and Definitions

Based on the types of vehicles individuals use, seven groups have been formed: Bus/Truck, Sedan/Small Truck, Motorcycle/Scooter, Bicycle, Pedestrian, Other vehicle, and Passenger/other people. This is to compute the fatal and injury rate within each group. Later, we shifted our focus to pedestrian-related data. Before transforming the data into individual-based and accident-based formats for further analysis, variable transformations would be applied to some features.

There are two outcomes in this study. A fatal accident is defined if there is any individual death within 24 hours or 30 days of the incident. Injury severity is defined based on the individual injury levels into four groups: death within 24 hours, death within 30 days, injury, and non-injury.

The definitions of key factors are referred to in [6] and some necessary adjustments according to our data. The definition of some specific variable transformation will be outlined. The cause of the accident is categorised into five groups, including three related to the driver: Improper driving at the pedestrian crossing, Not paying enough attention, or Others; one related to the pedestrian itself: mainly because crossing without following the regulations; and Unable to define.

The types of vehicles involved in the accident. If there is more than one type, it will be considered as Mix. Physical conditions include weather: Sunny, Cloudy, or Rainy; and light: Natural daylight, or Night/Early morning. Road types or conditions include road surfaces: Dry or Slippery; road types: Intersection, Straight Road, or Others; and the presence or absence of road edges.

Dates and times are categorized into seasons: Spring (March to May), Summer (June to August), Autumn (September to November), Winter (December to February); the day of the week: Weekday, or Weekend; and the time of the day: Midnight (12 am to 7 am), Morning Rush hour (7 am to 9 am), Midday (9 am to 5 pm), Afternoon Rush hour (5 pm to 7 pm), Evening and Night (7 pm to 12 am).

C. Driver's Information Extraction

Following the preparation of accident-based variables, we also need to extract the driver's information for the individual-based data. Variables include the number of drivers, driver's age (average for multiple drivers), sex (recorded as Mix if both are involved), proper driver's license, use of alcohol, and whether the driver hit or run after the incident. Merging the driver's information with pedestrian individual data creates an opportunity to explore the potential factors that are related to injury severity.

D. Outcome Analysis

To compare the overall traffic accidents and individuals by injury severity, Table I offers an overview of outcome distributions across different vehicle types.

Moving forward to pedestrian-based data, Table II employs the chi-squared test to evaluate the factors associated with fatal or non-fatal pedestrian-related accidents. Furthermore, using Logistic regression and focusing on p-values below 0.1 from the chi-squared test in Table II, adds a deeper layer of analysis to our understanding of the contributing factors. Although 5 factors meet the criteria, two of them (Light condition, and Hours of the day) were highly correlated with the time of the accident, so we only chose one for the subsequent analysis.

To find the high-risk periods for pedestrian-related accidents, heatmap analysis helps to gain a deep look. Fig. 1 and 2, revealed that there were more pedestrian-related accidents during the winter evenings (17 to 19 hours) compared to other time periods. The further exploration involved extracting the data in the 17 to 19 hours and dividing it into two groups, Spring/Summer and Autumn/Winter, to investigate which factors were contributing more accidents other than seasonal differences in Table III.

TABLE I. Distribution of Traffic Accidents and Individuals by Injury Severity.

Accidents Fatality		
	Fatal	Non-Fatal
Total Accidents (n, %)	262 (0.35%)	74967 (99.65%)
Vehicle Involved (n, %)		
Bus/Truck	33 (1.30%)	2496 (98.70%)
Sedan/Small Truck	149 (0.33%)	45229 (99.67%)
Motorcycle/Scooter	177 (0.26%)	66664 (99.74%)
Bicycle	23 (0.53%)	4283 (99.47%)
Pedestrian	101 (1.62%)	6150 (98.38%)
Other Vehicle	2 (0.61%)	327 (99.39%)
Passenger/Other People	26 (0.26%)	10118 (99.47%)

Individuals Injury Severity				
	Death 24 hours	Death 30 days	Injury	Non-injury
Total Individual (n, %)	193 (0.12%)	70 (0.04%)	99746 (60.91%)	63762 (38.93%)
Vehicle Used (n, %)				
Bus/Truck	0 (0.00%)	0 (0.00%)	48 (1.96%)	2402 (98.04%)
Sedan/Small Truck	6 (0.01%)	1 (0.00%)	2469 (5.09%)	46072 (94.90%)
Motorcycle/Scooter	93 (0.10%)	31 (0.03%)	77015 (84.84%)	13639 (15.02%)
Bicycle	11 (0.25%)	9 (0.20%)	3908 (88.44%)	491 (11.11%)
Pedestrian	73 (1.13%)	28 (0.43%)	6103 (94.17%)	277 (4.27%)
Other Vehicle	1 (0.31%)	0 (0.00%)	124 (38.63%)	196 (61.06%)
Passenger/Other People	9 (0.08%)	1 (0.01%)	10079 (93.55%)	685 (6.36%)

TABLE II. FACTORS IN FATAL OR NON-FATAL PEDESTRIAN-RELATED ACCIDENTS: CHI-SQUARED TEST AND LOGISTIC REGRESSION ANALYSIS

	Overall		Fatal		Non-fatal		Chi-squared	Multivariate Logistic Regression	
	n	%	n	%	n	%	p-value	OR (95% CI)	p-value
Sample size	6251	100.00%	101	1.62%	6150	98.38%			
Light Condition							0.049		
Natural daylight	3541	56.65%	47	1.33%	3494	98.67%			
Night or Early morning	2710	43.35%	54	1.99%	2656	98.01%			
Road Type							0.078		
Intersection	3597	57.54%	66	1.83%	3531	98.17%		ref	
Straight Road	2209	35.34%	33	1.49%	2176	98.51%		0.922 (0.590-1.441)	0.723
Other	445	7.12%	2	0.45%	443	99.55%		0.342 (0.083-1.413)	0.138
Road Edge Condition							0.092		
With Road Edge	1530	24.48%	17	1.11%	1513	98.89%		ref	
Without Road Edge	4721	75.52%	84	1.78%	4637	98.22%		1.543 (0.908-2.620)	0.109
Vehicles Involve							<.001		
Bus/Truck	129	2.11%	9	6.98%	120	93.02%		4.520 (2.134-9.573)	**<.001**
Sedan/Small Truck	3088	50.57%	51	1.65%	3037	98.35%		ref	
Motorcycle/Scooter	2478	40.58%	34	1.37%	2444	98.63%		0.858 (0.542-1.358)	0.514
Bicycle	209	3.42%	0	0.00%	209	100.00%		NA	1.000
Mix/Other	202	3.31%	7	3.47%	195	96.53%		2.359(1.052-5.452)	**0.037**
Hours of the Day							<.001		
Midnight	502	8.03%	31	6.18%	471	93.82%		5.972 (3.507-10.172)	**<.001**
Morning Rush hour	616	9.85%	11	1.79%	605	98.21%		1.675(0.824-3.408)	0.154
Midday	2425	38.79%	27	1.11%	2398	98.89%		ref	
Afternoon Rush hour	1111	17.77%	10	0.90%	1101	99.10%		0.870 (0.418-1.812)	0.710
Evening/Night	1597	25.55%	22	1.38%	1575	98.62%		1.252 (0.707-2.216)	0.441

E. Construction of Model

For the analysis of pedestrian injury severity, a Random Forest model will be used. First, the missing values should be addressed. Due to the target imbalance, undersampling would be applied to the injury group, while Synthetic Minority Oversampling Technique (SMOTE) methods would be applied to the two death groups. The input variables included key factors which are defined above, pedestrian's age, sex, and driver's information.

During the model training, 80 percent of the resampled dataset is designated for training, with the remaining 20 percent reserved for testing. Evaluation using 10-fold cross-validation to determine the best overall performance model. This ensures a robust evaluation of the model's predictive performance. Later, we validate the chosen model with the testing dataset, employing a confusion matrix to evaluate performance, and also output the feature importance for further insights.

F. Validation of Results

After constructing the Random Forest model, validation is crucial. The confusion matrix in Fig. 3 offers insights into the test results, allowing us to evaluate the model's performance and identify areas of misclassification.

In Fig. 4, the ranks of feature importance highlight the factors with the most significant roles in injury severity. Focusing on the top five features, we compare differences across the four groups, gaining valuable insights into the key factors that may influence pedestrian injury severity. This process not only identifies the contributing factors but also provides hints about how they extend their influence.

V. FINDING

In Table I, the overall analysis indicates ten times more fatal risks for the pedestrian compared to the overall average. This shows the urgency of pedestrian safety interventions.

Table II highlights the factors correlated with fatality including timing, location, and vehicle type. Logistic regression reveals higher risks in bus or complex accidents compared to others. Also, at midnight, the risk is much higher.

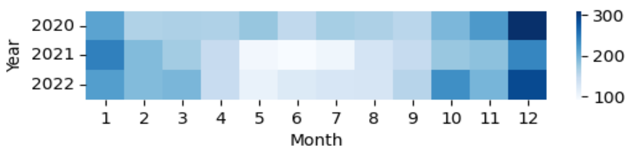

Fig.1 Monthly pedestrian-related accidents across three years (2020-2022).

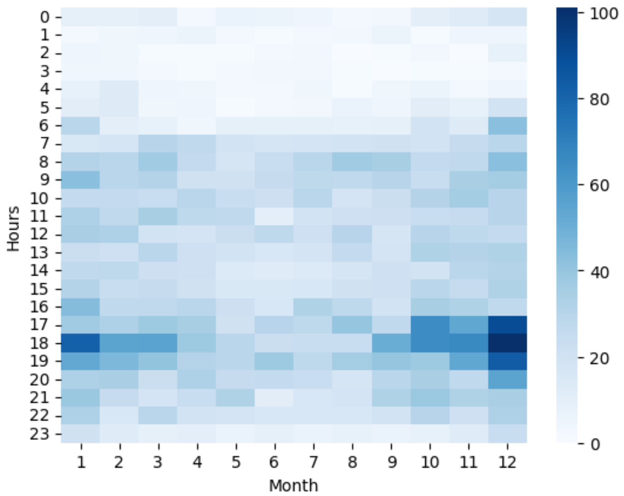

Fig. 2 Hourly distribution of accidents across months.

TABLE III. Comparison of Factors in Spring/Summer and Autumn/Winter Accidents During 17-19 Hours

	Overall (n, %)	Spring / Summer (n, %)	Autumn / Winter (n, %)	p-value
Sample size	1629	588 (36.10%)	1041 (63.90%)	
Weather				<.001
Sunny	747 (45.86%)	359 (48.06%)	388 (51.94%)	
Cloudy	312 (19.15%)	96 (30.77%)	216 (69.23%)	
Rainy	570 (34.99%)	133 (23.33%)	437 (76.67%)	
Light				<.001
Natural daylight	363 (22.28%)	227 (62.52%)	136 (37.47%)	
Night/Early morning	1266 (77.72%)	361 (28.52%)	905 (71.48%)	
Road surface condition				<.001
Dry	1028 (63.11%)	447 (43.48%)	581 (56.52%)	
Slippery	601 (36.89%)	141 (23.46%)	460 (76.54%)	
Vehicle Involve				<.001
Bus/Truck	18 (1.13%)	11 (61.11%)	7 (38.89%)	
Sedan/Small Truck	755 (47.34%)	270 (35.76%)	485 (64.24%)	
Motorcycle/Scooter	709 (44.45%)	238 (33.57%)	471 (66.43%)	
Bicycle	66 (4.14%)	40 (60.61%)	26 (39.39%)	
Mix/Other	47 (2.95%)	16 (34.03%)	31 (65.96%)	

Fig. 1 shows the number of pedestrian-related accidents within different months over three years. From May to July 2021, the number of accidents decreased. This is possible because, in May 2021, Taipei had a pandemic outbreak, although the city was not on lockdown, the government still suggested people stay home or work remotely, and this was not eased until July. However, the higher number in the winter still held across three years. Fig. 2 shows the number of accidents within different hours across months. And in the afternoon rush hour, more accidents were happening in winter than in summer.

Table III compares Spring/Summer and Autumn/Winter during 17 to 19 hours. It shows that weather, light, road surface, and vehicle involvement were the possible causes of these differences. The sunset time in winter is about 5 pm in Taiwan, and 7 pm in summer, with more rainy days causing the road surface to be slippery, and more cars and scooters involved. Combined with all the information, darker evenings, adverse weather, and rush hours might contribute to more pedestrian-related accidents that involve cars or scooters.

Fig. 3 shows the test results of the Random Forest model, the most misclassified are in the injury and non-injury groups. This pattern is also shown in the analysis of the top five important factors detailed in the later section. Fig. 4 shows the significance of age, ranking highest in importance and much higher than other factors. Displaying the distributions of these top factors within four groups provides insights into their contributions to injury severity.

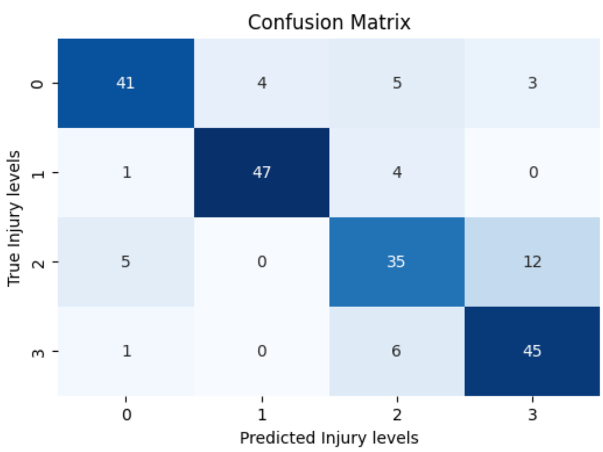

Fig. 3 Test results of the Random Forest Model.

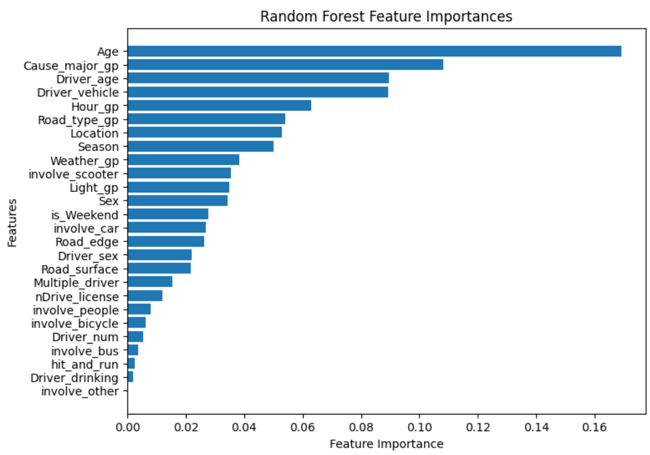

Fig. 4 Test results of the Random Forest Model.

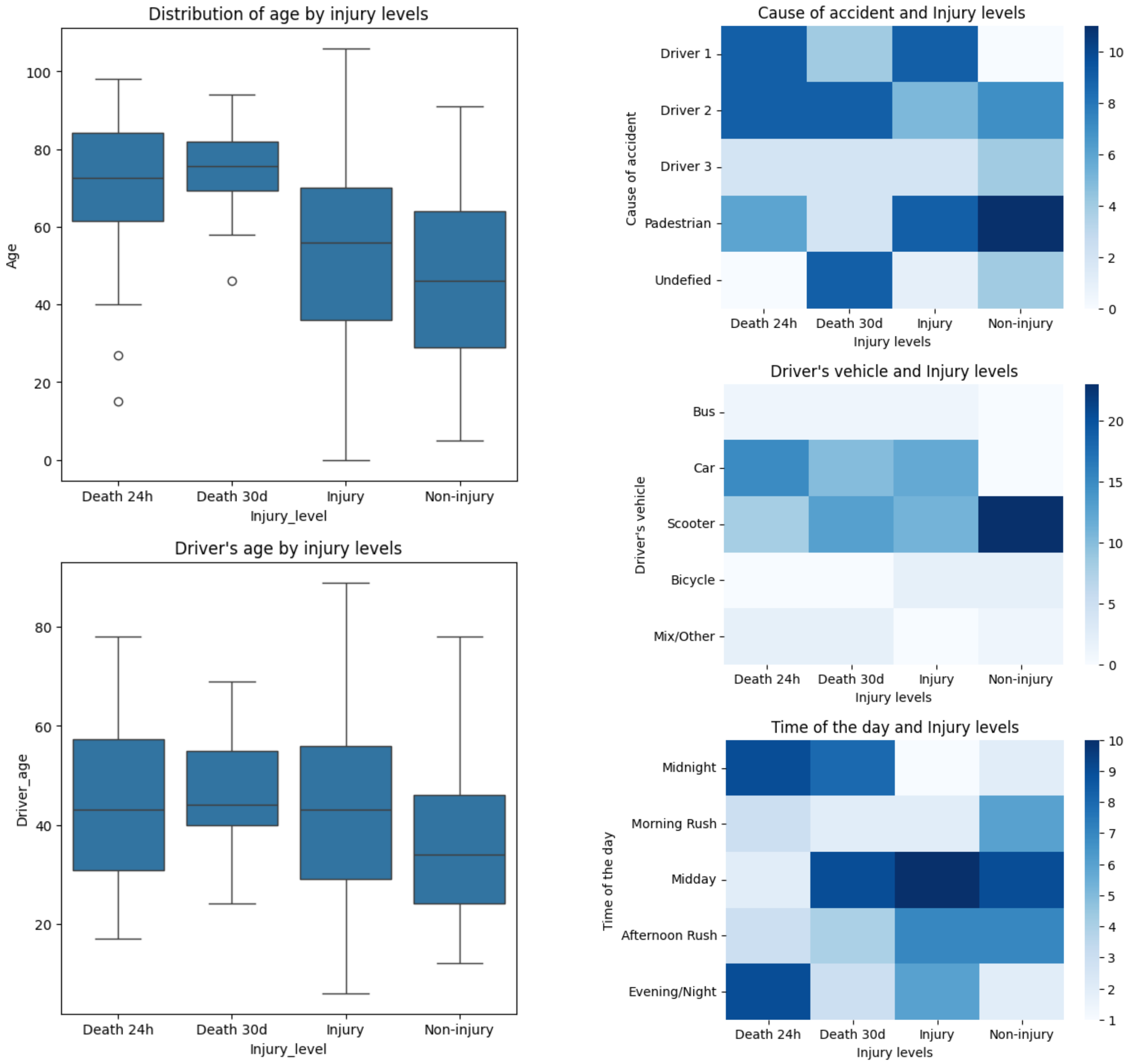

Fig. 5 Top five features across pedestrian injury severity.

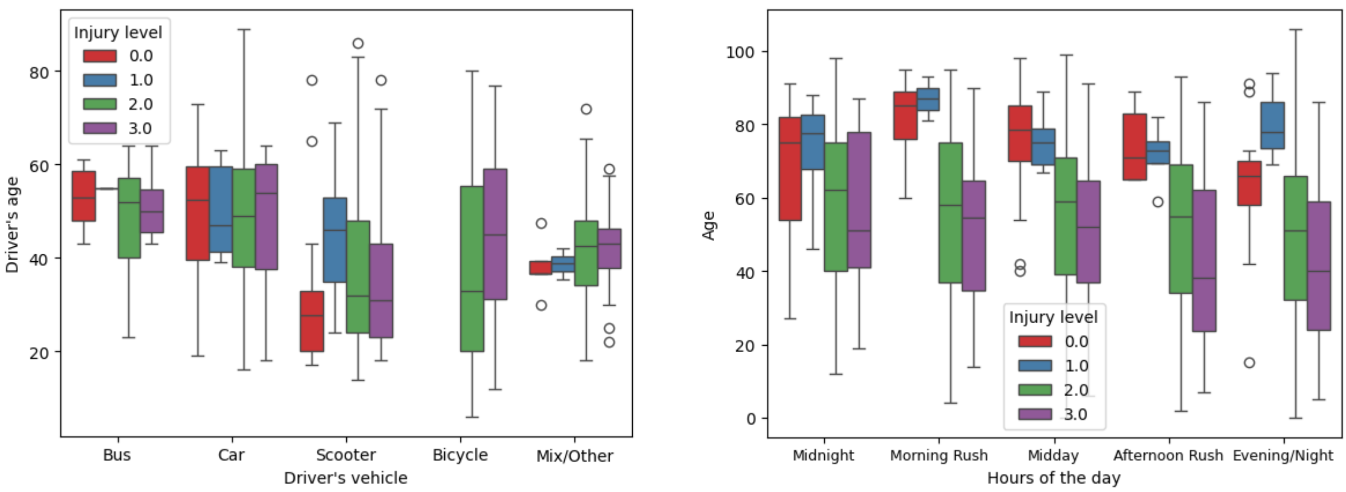

Fig. 6 Combining two factors across pedestrian injury severity.

Age shows a clear pattern, with older individuals more likely in death groups compared to injury or non-injury groups. The age is even higher in death within 30 days, suggesting there are still risks for older survivors after the initial 24 hours. Driver's age indicates a slightly younger age in the non-injury groups. Accidents caused by driver error tend to result in more severe outcomes. Vehicle involvement primarily included cars or scooters. Fatal accidents were more common at midnight. Combining factors unveils some interesting insights, such as the youngest scooter drivers involved in accidents death within 24 hours. Evening and midnight observations show the younger distributions in death groups, however, with death within 30 days it is much older than other groups in evening hours.

All these findings provide a robust foundation for regulatory actions and public warning. Including specific guidelines for evening rush hours in winter such as should be more careful in the evening rush hours in winter or encouraging people to wear colourful clothing during the darker light conditions [7]. Older generations can also be emphasised creating a safer environment for pedestrians.

VI. Reflections and Further Work

The findings are consistent in accidents and individual results and have similar patterns in [6] and [8], showing the reliability of the findings. However, it is crucial to note potential bias from data recorded during the COVID-19 pandemic. Further investigations into the interaction of these factors would enhance the study's depth. Future research could categorise injured pedestrians based on their injured body parts, analyse geographic location, and include more data from 2012 to 2019 increasing the sample size, offering more information and perspectives for answering the analytical questions and providing more perspective to enhance pedestrian safety.

Word Counts	Max	Counts
Abstract	150	146
Introduction	300	228
Analytical questions	300	199
Data (Materials)	300	217
Analysis	1000	882
Findings, reflections and further work	600	602

References

[1] W. Chang, 'Taiwan's "living hell" traffic is a tourism problem, say critics', *CNN*, Dec. 06, 2022. Accessed: Dec. 08, 2023. [Online]. Available: https://edition.cnn.com/travel/article/taiwan-traffic-war-tourism-intl-hnk/index.html

[2] 'Taiwan Country Security Report', Overseas Security Advisory Council, Bureau of Diplomatic Security, U.S. Department of State. Accessed: Dec. 08, 2023. [Online]. Available: https://www.osac.gov/Content/Report/a226918f-511c-4549-971b-1ccf3f047af2

[3] '111 年全國行人件數各年齡層分布 (Distribution of pedestrian-related traffic accidents nationwide in 2022.)', Ministry of Transportation and Communications, R.O.C. Accessed: Dec. 08, 2023. [Online]. Available: https://roadsafety.tw/AccOrder?Order=Age&type=%E8%A1%8C%E4%BA%BA

[4] K. E. Charters, B. J. Gabbe, and B. Mitra, 'Population incidence of pedestrian traffic injury in high-income countries: A systematic review', *Injury*, vol. 48, no. 7, pp. 1331–1338, 2017, doi: https://doi.org/10.1016/j.injury.2017.05.021.

[5] 'A1 & A2 types of traffic accidents in Taipei', Ministry of Digital Affairs, R.O.C. Accessed: Dec. 08, 2023. [Online]. Available: https://data.gov.tw/en/datasets/130110

[6] C.-W. Pai, P.-L. Chen, S.-T. Ma, S.-H. Wu, V. Linkov, and H.-P. Ma, 'Walking against or with traffic? Evaluating pedestrian fatalities and head injuries in Taiwan', *BMC Public Health*, vol. 19, no. 1, p. 1280, 2019, doi: 10.1186/s12889-019-7588-1.

[7] World Health Organization and others, 'PEDESTRIAN SAFETY A road safety manual for decision-makers and practitioners', 2023.

[8] H.-P. Ma, P.-L. Chen, S.-K. Chen, L.-H. Chen, V. Linkov, and C.-W. Pai, 'Population-based casecontrol study of the effect of sun glare on pedestrian fatalities in Taiwan', *BMJ Open*, vol. 9, no. 8, 2019, doi: 10.1136/bmjopen-2018-028350.