ARCHIVES OF TRANSPORT
Volume 68, Issue 4, 2023

ISSN (print): *0866-9546*
e-ISSN (online): *2300-8830*
DOI: *10.61089/aot2023.ttb8p367*

ANALYSIS OF VEHICLE PEDESTRIAN CRASH SEVERITY USING ADVANCED MACHINE LEARNING TECHNIQUES

Siyab UL ARIFEEN[1], Mujahid ALI[2], Elżbieta MACIOSZEK[3]

[1] *Department of Civil Engineering, COMSATS University Islamabad, Abbottabad, Pakistan*

[2, 3] *Department of Transport Systems, Traffic Engineering and Logistics, Faculty of Transport and Aviation Engineering, Silesian University of Technology, Katowice*

Abstract:

In 2015, over 17% of pedestrians were killed during vehicle crashes in Hong Kong while it raised to 18% from 2017 to 2019 and expected to be 25% in the upcoming decade. In Hong Kong, buses and the metro are used for 89% of trips, and walking has traditionally been the primary way to use public transportation. This susceptibility of pedestrians to road crashes conflicts with sustainable transportation objectives. Most studies on crash severity ignored the severity correlations between pedestrian-vehicle units engaged in the same impacts. The estimates of the factor effects will be skewed in models that do not consider these within-crash correlations. Pedestrians made up 17% of the 20,381 traffic fatalities in which 66% of the fatalities on the highways were pedestrians. The motivation of this study is to examine the elements that pedestrian injuries on highways and build on safety for these endangered users. A traditional statistical model's ability to handle misfits, missing or noisy data, and strict presumptions has been questioned. The reasons for pedestrian injuries are typically explained using these models. To overcome these constraints, this study used a sophisticated machine learning technique called a Bayesian neural network (BNN), which combines the benefits of neural networks and Bayesian theory. The best construction model out of several constructed models was finally selected. It was discovered that the BNN model outperformed other machine learning techniques like K-Nearest Neighbors, a conventional neural network (NN), and a random forest (RF) model in terms of performance and predictions. The study also discovered that the time and circumstances of the accident and meteorological features were critical and significantly enhanced model performance when incorporated as input. To minimize the number of pedestrian fatalities due to traffic accidents, this research anticipates employing machine learning (ML) techniques. Besides, this study sets the framework for applying machine learning techniques to reduce the number of pedestrian fatalities brought on by auto accidents.

Keywords: *Machine learning, ANN, BNN, Vehicle-pedestrian crash*

To cite this article:

Ul Arifeen, S., Ali, M., Macioszek, E., (2023). Analysis of vehicle pedestrian crash severity using advanced machine learning techniques. Archives of Transport, 68(4), 91-116. DOI: https://doi.org/10.61089/aot2023.ttb8p367

Contact:
1) siyabularifeen@cuiatd.edu.pk [https://orcid.org/0009-0001-3785-9404] 2) mali@polsl.pl [https://orcid.org/0000-0003-4376-0459] – corresponding author; 3) elzbieta.macioszek@polsl.pl [https://orcid.org/0000-0002-1345-0022]

1. Introduction

Walking is a vital part of daily transportation since it is a physically and environmentally healthy mode of travel, especially for first and last-mile movement. However, due to their lack of protection, pedestrians or foot travelers are the most venerable road users who are more likely to sustain severe injuries or die in car accidents (Behnood and Mannering, 2016, Pucher and Dijkstra, 2003, Liu et al., 2019, Hafeez et al., 2023). Significant socioeconomic repercussions result from pedestrian mortality and injuries in traffic accidents. This is especially important when taking into account the ongoing initiatives taken by developed economic countries to increase road safety. Since almost anyone can use the road, pedestrians are the prevailing road user. From 2017 to 2019, vehicle-pedestrian collisions accounted for 18% of all traffic accidents in Hong Kong (Hong Kong Transportation Department, 2022, July 27, Zhu, 2022). Numerous studies on pedestrian injuries have been conducted to understand the features that consequentially affect the seriousness of vehicle-pedestrian collisions. In Hong Kong, buses and the metro are used for 89% of trips, and walking has traditionally been the primary way to use public transportation. Accessibility, connectedness, environment, and safety can be used to describe a city's walkability. In Hong Kong, pedestrians made up 17% of the 20,381 traffic fatalities in 2015 (Xing et al., 2019). However, 66% of the fatalities on the highways were pedestrians. People choose walking by foot for several purposes, such as recreation, getting to and from work, school, or local businesses, and connecting with other modes of transportation. In the National Road Safety Strategy, foot travelers are categorized as vulnerable road users. They have a fragile defense in crashes compared to other road users (Australian Transport Council, 2011, May 20). Despite the decline in pedestrian fatalities resulting from traffic accidents in Hong Kong, researchers have been on the lookout for opportunities to grasp more knowledge about the factors that affect crash probability to accurately predict the likelihood of foot travelers involved in crashes and use that information to guide policy initiatives and prevention strategies to minimize the frequency of pedestrian crashes. However, it is crucial to pinpoint the causes of the greater risk of injury and crash for pedestrians. Numerous researches have shown that the environmental conditions around roads, speed of vehicles, crash conditions, traffic management, and pedestrian and driver behavior all have an impact on pedestrian safety (Sze and Wong, 2007, Gårder and Prevention, 2004, Jian et al., 2005, Zajac, 2003, Aziz et al., 2013, de Lavalette et al., 2009, Dai, 2012, Khuzan and Al-Jumaili, Luke, 2023, Wang et al., 2022).

As an active means of transportation, walking is vulnerable in inclement weather conditions. Due to the escalating differences in air temperature and precipitation during climate breakdown, inclement weather phenomena, including storms, heat waves, and high precipitation, are happening more regularly. One of the fundamental environmental features associated with pedestrian accidents has been determined to be the weather. Kim., et al. and Vilaça, N. Silva., et al. (Kim et al., 2010, Vilaça et al., 2017, Maze et al., 2006) discovered that, in good weather, the collision likelihood of pedestrians is 77% lower than that of motorcyclists. According to reports (Mohamed et al., 2013, Maze et al., 2006, Theofilatos and Yannis, 2014), higher rainfall frequency and severity are related to lower pavement friction, poor visibility, and increased risk of crashes. Additionally, the relationship between the seriousness of a pedestrian injury and other risk variables may interact depending on the weather. Li, D. et al. (Li et al., 2017) observed that under favorable weather conditions, light conditions might influence how severe a pedestrian injury is, but no evidence of a link in unfavorable weather conditions was found. Therefore, it is crucial to investigate how weather and its extremes affect the likelihood of pedestrian injury.

Since possible dissensions are created by the movement of different sorts of road users, intersections represent a significant proportion of all vehicle-pedestrian collisions. The complexity of conflicting movements may also make intersection collisions more serious. The severity analysis of vehicle crashes with foot travelers at intersections, however, has received relatively little attention (Abdel-Aty et al., 2005). A significant fraction of traffic accidents occur at signalized intersections, which are dangerous location type on the road (Ali et al., 2023, Marzoug et al., 2022). Crash regularity and severity are the two primary considerations in determining the correlation between crash occurrences and different risk factors to build cost-effective safety remedies. On the one hand, many researchers have examined the crash frequency at intersections for various crash categories with a focus on crash prediction models (Kim and Washington, 2006, Kim et al., 2006). However, the intensity of crashes is another issue for traffic safety. The level of pedestrian safety depends on the availability of specialized facilities for pedestrians. A thorough understanding of how risk factors affect crash severity would be immensely useful before creating and implementing traffic safety interventions.

There are various methods for analyzing crash severity for several reasons. The majority of earlier severity investigations have often used categorical data analysis methodologies. Some researchers used binomial/multinomial logistic models (Mercier et al., 1997, Mannering and Grodsky, 1995, Shankar and Mannering, 1996, James and Kim, 1996, Hilakivi et al., 1989, Al-Ghamdi, 2002) considering crash potential as a nominal variable, to investigate the significance of risk factors, whereas another study demonstrates the use of ordered logit/probit models (Abdel-Aty et al., 2005, Rifaat and Chin, 2005, Quddus et al., 2002, O'donnell and Connor, 1996), which take severity level ordering into account. The great majority of crash severity studies used statistical regression models, which may not perform well when driving large amounts of sophisticated accident data with lots of discrete variables or variables with numerous categories. Furthermore, the linearity assumption, which is widely used in these models, is difficult to validate in crash cases (Ding et al., 2018, Li et al., 2018).

In comparison to conventional statistical models, ML techniques are more adaptable since they can investigate outliers, turbulent and missing data while making no or few prior presuppositions about the inputs (Aghaabbasi et al., 2020, Aghaabbasi et al., 2021, Chen et al., 2021). Another well-known example of data-driven strategies that are designed to increase the effectiveness and accuracy of accident data rectification and forecasting is the use of ML techniques. Early studies used a range of ML approaches, including decision trees (DT) ensemble learning, support vector machines (SVM), and artificial neural networks (ANN), to predict the seriousness and frequency of pedestrian crashes (Nayeem et al., Zhao et al., 2023, Tao et al., 2022). Their outcomes show that these strategies can perform better than conventional approaches and are very adaptable. Consequently, this study chose to analyze data related to vehicle-pedestrian crashes in Hong Kong from 2016-2018 using a BNN based on machine learning. The use of Bayesian computing techniques is growing in popularity as a result of improvements in computing techniques. The advantage of using Bayesian models is that they can handle exceedingly complex models, especially those with complex probability functions. However, ordinary NN models are now heavily criticized for their inability to adequately match training data and for producing anticipated results with unfavorable variances (Xie et al., 2007, Ali et al., 2021). One of the main reasons for this problem is overfitting. The fundamental NN model is more capable of linear and nonlinear estimating than typical statistical approaches, however, this method's generalizability is constrained by its susceptibility to the overfitting problem, which reduces its usefulness for predicting accident severity and frequency. Applying the Bayesian approach to NN models has been proven (Liu et al., 2020, Marzban and Witt, 2001, Yan et al., 2017) to greatly reduce overfitting while retaining the NN's superior nonlinear approximation capability. However, the aforementioned combination has just rarely been implemented in the field of crash prediction, particularly for vehicle-pedestrian crashes.

The primary objective of this research is to evaluate the efficacy of the BNN model in predicting the vehicle-pedestrain crash severity. In addition, our research makes a scholarly contribution to the field of pedestrian road crash fatality modeling in the subsequent manners: (1) constructing a composite of architectures to evaluate the efficacy of the model; (2) assessing a range of attributes that may aid in the classification and prediction of pedestrian injuries; (3) comparing the performance of BNN with other machine learning models.

2. Literature review

The research of Yasmin, Naveen, et al. (Yasmin and Eluru, 2013), contrasts various approaches to simulating the extent of injuries caused by drivers in vehicle accidents. Ordered logit, multinomial logit, generalized ordered logit, nested logit, mixed generalized ordered logit, mixed multinomial logit, and ordered generalized extreme value logit are examples of these techniques, which are also referred to as ordered and unordered response models. The study exploits data from the 2010 General Estimates

System, a national database of American traffic accidents, to evaluate the efficacy of these models using several different metrics. The analysis's findings indicate that when it comes to estimating the intensity of driver injuries, the mixed generalized ordered logit model competes fiercely with the mixed multinomial logit model. The study also investigates the impact of potential under porting on these models, both with and without corrections of data. The study by Zhou, Z. Ping, et al. (Zhou et al., 2013) examines various pedestrian behavior at signalized crossings in China, a country with a peak frequency of pedestrian accidents. The study involved watching how pedestrians behaved in the real world and then giving the same people a questionnaire to learn more about their demographics, socioeconomic level, attitudes, and preferences. In Nanjing, surveys of 1878 pedestrians at 16 crossings were conducted. The research team then created a multinomial logit model to comprehend how these variables affect pedestrian behavior after doing correlation analysis and structural equation modeling on the data to find five latent variables (flexibility, conformity, fastness, safety, and comfort). The study's findings indicate that age and gender are the most significant determinants for full sneakers, whereas approaching time, the existence of approaching autos, and the length of the crosswalk are the most significant factors in predicting whether a pedestrian would be a late beginning. Accidents between people and vehicles frequently cause serious injuries or fatalities. This study by Chen, and Z.Fan (Chen, 2019) employed mixed logit models to study the variables that affect how seriously injured pedestrians are in these kinds of collisions in both countryside and inner-city areas of North Carolina, the United States. The Highway Safety Information System database was used to glean data on pedestrian-vehicle collisions that occurred between 2005 and 2012 for the study. Five categories were used to categorize the severity of the injuries sustained in these collisions: fatalities, incapacitating injuries, obvious injuries, potential injuries, and no injuries (just property damage). The study's findings demonstrated that several elements, including the driver's physical condition, the presence of large trucks, dim lighting, and specific speed limits, greatly affect pedestrian injuries in both rural and urban locations. The study's results can be utilized to help create measures to lessen the seriousness of the injuries suffered by pedestrians in collisions with vehicles and enhance traffic safety.

The purpose of the study regulated by Das et al. (Das et al., 2020) aimed to provide a system for classifying pedestrian and bicycle crash types from unstructured data, such as content from police reports, using machine learning. The researchers collected data from two localities in Texas and found that the XGBoost model was the most accurate at classifying crash types, with an accuracy rate of up to 77% for training data and 72% for test data. This shows that ML is capable of recognizing patterns and trends in crash processes involving pedestrians and bicycles, which are frequently not fully documented in state or national databases. The results serve as a foundation for utilizing machine learning to increase the effectiveness and precision of the Pedestrian and Bicycle Crash Analysis Tool (PBCAT), a tool used to examine collisions involving walkers and bikes. The purpose of the study carried out by Rahimi et al. (Rahimi et al., 2020) aimed to discover trends in fatal pedestrian and bicycle incidents in Florida involving heavy vehicles. The study used decision trees and random forests, two machine learning algorithms, with data from 2007 to 2016. The findings showed that depending on the volume of traffic and kind of route, different factors affected the severity of the crash and the possibility of fatalities. The proximity to an intersection and the speed of the vehicle were significant features in examining the severity of the incident on local roads with large traffic volumes (more than 38,000 annual average daily traffic). Accidents involving young or middle-aged truck drivers on divisional roads with speed limits between 50 and 75 mph that happened in the middle of the day were more likely to be deadly than those involving older truck drivers on low-volume, local routes. When the weather was clear, median barriers and curb shoulders near signalized intersections were linked to higher fatality rates on low-volume highways. According to Guo et al. (Guo et al., 2021) for traffic agencies, the safety of pedestrians, especially those over 65, is a top priority. Older pedestrians are at a high likelihood of being injured or dying when they are involved in traffic accidents. It's critical to carefully examine the elements that contribute to collisions involving elderly pedestrians in order to better protect this demographic. In this work, the severity of aged road walker traffic crashes in

Colorado, US, was classified using a machine learning model called Extreme XGBoost. According to Shapley Additive explanations (SHAP), which were used to explain the model's findings, the driver's attributes, the characteristics of elder pedestrians, and vehicle movement were the main variables determining the likelihood of the three different crash severity levels. By addressing these crucial elements, this information can benefit traffic management and road infrastructure agencies in protecting senior pedestrians. The study of Saha (Saha and Dumbaugh, 2021) examines, the associations between the incidence of pedestrian crashes and other elements of the existing environment at the level of census block groups. The authors analyze data from Broward and Miami-Dade Counties in Florida, including land use, traffic parameters, road network, and sociodemographic using an ML technique known as the component-wise model-based gradient boosting algorithm. The approach supports the use of a variety of base learners, including Markov Random Fields, generalized additive models, and decision trees. Decision trees perform worse than generalized additive models in the study's performance comparison of various base learners. Urban planners and decision-makers can use the study's findings, which demonstrate both linear and non-linear correlations between a few variables and the frequency of pedestrian crashes, to enhance pedestrian safety. The goal of the study conducted by Zhu. (Zhu, 2022) is to comprehend the elements that affect serious pedestrian-vehicle incidents at crossings. The research applies data from Hong Kong that spans three years from 2016-2018 and analyses the data using a variety of data mining approaches, including CART, GB, RF, ANN, and SVM models. For comparison, the logistic regression model, which has been widely applied in earlier studies, is considered a baseline. The findings suggest that when there is slight rain or dazzling and the intersection is regulated by a traffic signal or has no control, the likelihood of a fatal or serious pedestrian-vehicle collision increases. On the other hand, when the climate is clear, the light is either daylight or dark, and in key locations in Hong Kong, the severity of the accident tends to reduce. The authors offer policy recommendations and preventative strategies based on these findings to lower the incidence of fatal and serious pedestrian-vehicle collisions at crossings. The study of Zhao et al. (Zhao et al., 2023) aims to examined the occurrence and severity of pedestrian collisions throughout the state of Texas by employing tree-based ML models. A comprehensive analysis was conducted utilizing a decade's worth of police records, in conjunction with roadway inventory and supplementary data, to accurately delineate the geographical distribution of over 78,000 pedestrian collisions across more than 700,000 distinct road segments. In the study, various methodologies such as Bayesian additive regression trees (XBART), random forests (RF), and gradient boosting (XGBoost, LightGBM) are employed and subsequently compared. All four specifications have comparable performance in forecasting crash counts, however LightGBM exhibits significantly quicker computational speeds. The Lee et al. (Lee et al., 2023) utilized data pertaining to pedestrian traffic incidents, which were categorized into three distinct levels of injury severity: modest, severe, and deadly. In order to conduct an investigation on the performance of several models, XGBoost, logistic regression, CatBoost, Naïve Bayes, and LightGBM were employed. The analysis involved the use of five ML methods, and hyperparameter tuning was conducted to enhance the model's efficiency. In this investigation, the classification accuracies of the five ML models were recorded as 0.705, 0.688, 0.708, 0.577, and 0.707, respectively. Among these models, LightGBM had the highest classification accuracy of 0.708. The primary aim of the research carried out by Lu et al. (Lu et al., 2022) was to establish an analytical structure that effectively combines ML techniques with route analysis in order to quantitatively assess behavioral pathways in bicycle-motor collisions with cars. The study investigated five distinct ML techniques, including RF, SVM, Categorical Naive Bayes (CNB), AdaBoost, and NN. In order to mitigate the potential bias inherent in individual models, the article presents a methodology that suggests the aggregation of model estimates through the process of averaging marginal effects. The present study utilized a dataset of 9,296 incidents involving collisions between bicycles and motor vehicles in order to illustrate the implementation of the proposed framework. The findings from the analysis of five distinct ML models indicate a consensus regarding the direction of marginal effects. However, notable disparities are observed in the magnitudes of these effects. In their work Yang et al. (Yang et al., 2022), three widely recognized

ML models were utilized to forecast pedestrian mortality associated with traffic crashes. The models utilized in this study encompassed k-nearest neighbors (KNN), SVM, and ensemble decision trees (EDT). The models underwent hybridization using a Bayesian optimization (BO) algorithm in order to identify the optimal values of their hyperparameters. The results of this study indicate that the performance of all three models was enhanced through the utilization of BO. The KNN model demonstrated the most substantial enhancement in accuracy, with an increase of 11%, following the application of the BO technique. Nevertheless, the SVM and EDT models exhibited superior accuracy compared to the KNN model.

3. Data

The dataset for model estimation for vehicle-pedestrian crashes derives from the Hong Kong traffic department (Hong Kong Transportation Department, 2022, July 27) which includes 9090 records of vehicle-pedestrian crashes from 2016 to 2018 with complete data entries. The severity degree of the vehicle-pedestrian crash is the target variable, and predictors taken into account include time factors, vehicle movement, environmental circumstances, etc. These factors enabled us to complete the study's goal, which was to analyze pedestrian crash data using a neural network and Bayesian theory. This study can be expanded in the future by using datasets with more factors like driver's age (Mercier et al., 1997, Al Mamlook et al., 2020), sex (Mafi et al., 2018), time of accident (Olowosegun et al., 2022), and detailed weather conditions (Mondal et al., 2020, Liu et al., 2021). It is important to note that the following transformations and normalizations were performed on the input dataset:

Rearranging the nominal variable's order resulted in the largest category coming last and the smallest category showing first. Missing values were replaced with the mean in continuous variables. Missing values in nominal variables were replaced by the mode. For missing values in ordinal data, the median was used as a replacement. Table 1 illustrates the parameters of the variables.

4. Methodology

A flowchart outlining the general procedure for determining the severity of the vehicle-pedestrian accident in this paper is shown in Figure 1. To accomplish the goal of this study, the BNN algorithm was used. More in-depth explanations of the stages are provided in the following sections.

4.1. Theory of BNN

To predict and categorize vehicle-pedestrian crash severity, this study used a Bayesian approach. Bayesian models make use of Bayes' theorem to extract and specify characteristics of a likelihood distribution from data collected in Equation (1).

$$P\,(X \mid Y) = \frac{P\,(X \mid Y)}{P\,(Y)} \tag{1}$$

P(X|Y) is the probability of event X occurring given that event Y has already occurred. P(Y|X) is the probability of event Y occurring given that event X has already occurred, and P(X) and P(Y) are the marginal probabilities of events X and Y respectively. Bayes' theory provides a way to estimate the probability of an event by taking into account not only past data, but also the "prior" probability of the event.

Assuming that the network's variables are arbitrary and dependent on an earlier probability distribution, a BNN is a NN that is designed to fit actual figures (Neal, 2012, Khumukcham et al., 2023). During the training stage, various NN types use a variety of strategies to learn from the input and change the network weights (Wang and Yeung, 2016). A standard NN may approximate a single data point after training, and its weights are thought to be deterministic. The weights of the BNN are displayed as likelihood distributions across potential data points rather than assuming a single-point estimation following training. The weights network distribution variance reflects the BNN's performance uncertainty (Yaseen et al.). Figure 2 illustrates the difference between a BNN and a stochastic NN.

For the binary classification between the two assignments, the authors used a BNN in this study: 0 for a low-severity crash and 1 for a severe or deadly vehicle-pedestrian crash while taking data uncertainty into account. The authors train the BNN, by using an optimization technique, like variational inference (VI) for estimating likelihood densities. Contrary to more traditional approaches like the Markov chain and Monte Carlo, VI creates the variables of these distributions instead of the weights directly.

Table 1. Hong Kong vehicle-pedestrian collision dataset descriptive statistics

Variable	Description	Count	Fatal	Serious	Slight
Severity	Fatal accident	200	—	—	—
	Serious injury accident	1944	—	—	—
	Slight injury accident	6946	—	—	—
Junction Control	No control	1291	24	243	1024
	Stop (halt)	71	2	14	55
	Give way (slow)	400	7	71	322
	Traffic signal	1301	54	493	954
	Police	2	0	0	2
	Not junction	6024	113	1323	4588
Weather	Clear	8634	183	1838	6613
	Dull	358	13	83	262
	Fog/mist	65	4	19	42
	Strong wind	23	0	4	19
Rain	Not raining	8071	168	1708	6195
	Light raining	830	30	186	614
	Heavy raining	167	2	47	118
Year	2016	3172	81	734	2357
	2017	2958	60	635	2263
	2018	2960	59	575	2326
Light Condition	Daylight	6512	136	1394	4982
	Dawn/Dusk	372	12	84	276
	Dark	2196	52	466	1678
Hit And Run	No	8772	194	1892	6686
	Yes	318	6	52	260
Vehicle Movement	1 moving vehicle	9001	189	1908	6904
	2 moving vehicle-from the same direction	64	6	23	35
	2 moving vehicles - from the opposite direction	7	1	3	3
	2 moving vehicles - from different roads	8	1	4	3
	> 2 moving vehicles - from the same direction	9	3	5	1
	> 2 moving vehicles - from the opposite direction	0	0	0	0
	> 2 moving vehicles - from different roads	1	0	1	0
District	Central and Western	649	10	145	494
	Eastern	543	11	178	354
	Islands	189	2	37	150
	Kowloon City	592	18	96	478
	Kwun Tong	699	18	172	509
	Kwai Tsing	474	13	94	367
	Yau Tsim Mong	1198	27	170	1001
	North	365	9	86	270
	Southern	313	4	85	224
	Sai Kung	244	8	61	175
	Sham Shui Po	709	21	95	493
	Sha Tin	425	8	67	350
	Tuen Mun	318	3	87	228
	Tai Po	311	7	72	232
	Tsuen Wan	418	16	69	333
	Wan Chai	708	5	140	563
	Wong Tai Sin	386	6	106	274
	Yuen Long	649	14	184	451

It is possible to consider the BNN utilized in this research as a probabilistic model $P(b \mid a, \gamma)$. Here, 'b' represents a grouping of our classification (b = 0 or 1); 'a' represents a set of attributes, $'\gamma'$ represents the weight parameter; and $P(b \mid a, \gamma)$ represents a categorical probability. The training dataset K could be used to produce the likelihood function (LF) that depends on the parameter Y. The LF is as follows:

$$P(K \mid \gamma) = \Pi P(b \mid a, \gamma) \tag{2}$$

By maximizing the LF with negative log-likelihood as the objective function, can derive the maximum likelihood estimate (MLE). The posterior distribution, according to the Bayes theory, is proportional to the result of the prior distribution, P(γ), and the probability P(K| γ). Alternatively, MLE uses point computation for the parameters; as a result, the weight's unreliability is not taken into account. A BNN model, therefore, averages predictions from several NN models that are weighted as per the posterior distribution of the γ. The posterior predictive distribution's mathematical equation is as follows:

$$P(b \mid a, K) = \int P(b \mid a, \gamma)\, P(\gamma \mid K) d\gamma \tag{3}$$

4.2. Theory of Artificial Neural Network

As concluded from Figure 3, the ANN technique is based on a configuration of interconnected nodes and weighted linkages. The algorithm is based on how the human brain processes information from data, learns from it, and then categorizes it in response (Şen, 2023, Arifeen et al., 2023). To identify distinctive patterns, artificial neural networks resolve the complicated relationship between input and output.

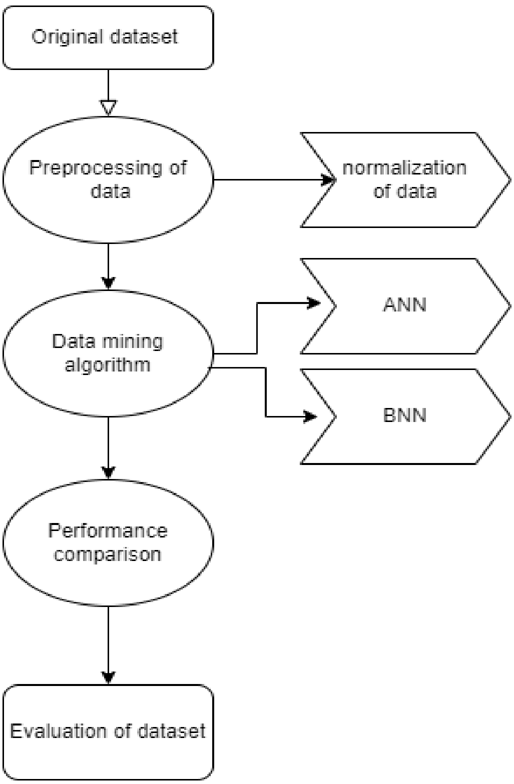

Fig. 1. Flowchart of the study

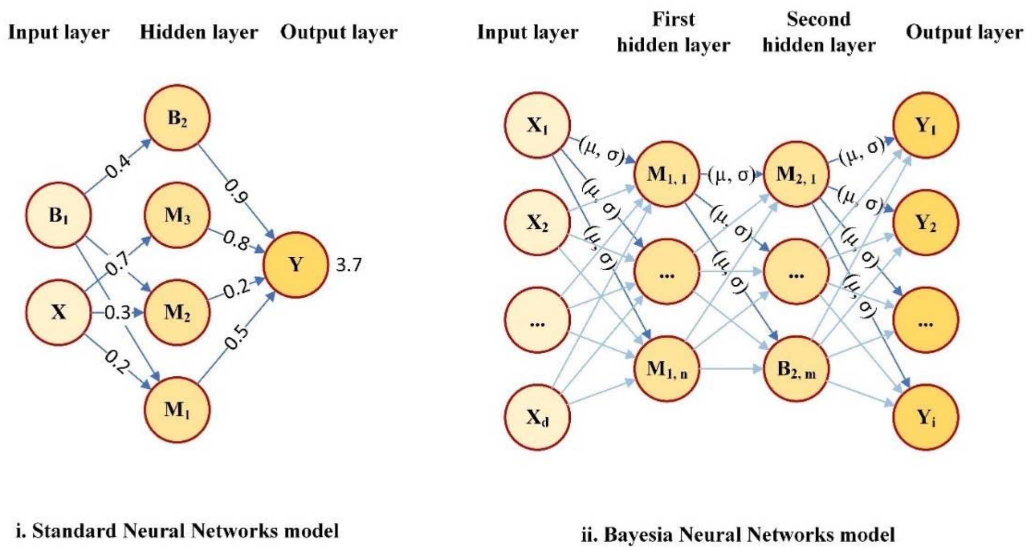

Fig. 2. Standard structures of NN and BN (Tao et al., 2022)

The input layer in this study collects numerical data on vehicle-pedestrian collisions. The incoming input is then subjected to several mathematical operations by the hidden layer to discover data patterns. After giving each input node a weight, the bias is calculated using a transfer function along with the weighted total of the inputs. The findings of the prediction are also compared to a specified threshold that has been established for classifying the intensity of vehicle-pedestrian crashes using the activation function. The learning process(Chen et al., 2017, Zhu, 2022) changes the weight values to make the ANN model's output compatible with the class tabs in the training set.

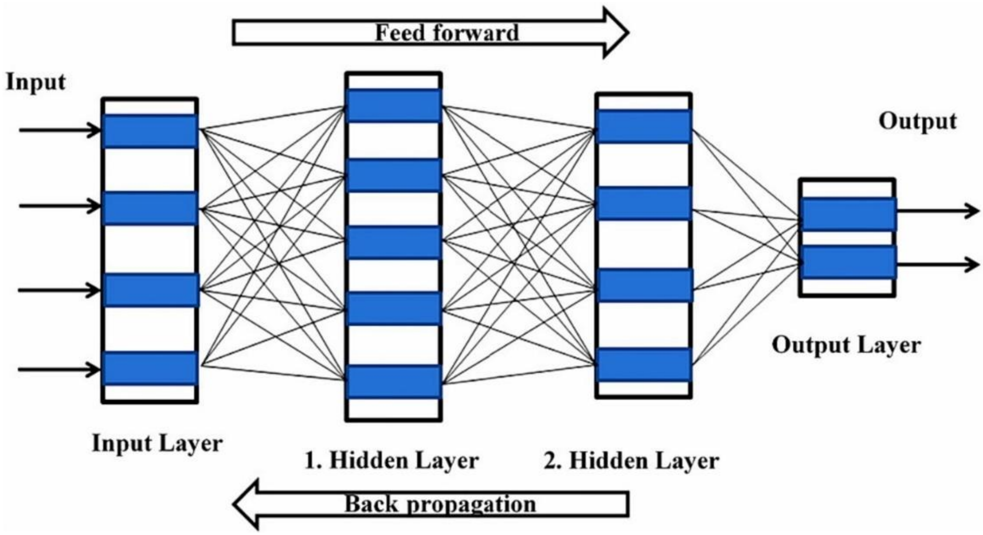

Fig. 3. Typical structure of ANN (Arifeen et al., 2023)

5. Evaluation of Model Performance

The BNN model's accuracy in correlating and predicting pedestrian mortalities and injuries as a result of traffic accidents was assessed using the following criteria.

5.1. The area under the ROC curve (AUC)

A binary classifier's accuracy is measured by the area under the Receiver Operating Characteristic (ROC) curve. It is utilized to evaluate the performance of several models and choose the best threshold for a particular model. The two processes required to calculate the area under the ROC curve are plotting the true positive rate vs the false positive rate for various thresholds and computing the AUC. The more distinct the model is between the two classes, the bigger the AUC value.
This study estimated a scoring classifier at various cutoffs using the AUC value. A model's capacity to discriminate between positive and negative classifications is measured by the AUC.

5.2. Matthew's correlation coefficient (MCC)

The effectiveness of binary classifications is gauged by the Matthews Correlation Coefficient (MCC). It serves as a gauge for how well a model predicts positive and negative classes. It is computed by adding up all of the true positives, true negatives, false positives, and false negatives, then multiplying that total by the square root of the product of the true positives, false negatives, true positives, and false negatives, respectively.
In the present research, the MCC was used to gauge the accuracy of binary categorizations. The MCC takes into account both reliable and unreliable positives and negatives, making it a stabilized measure that can be used even when the categories have vastly different sizes. This criterion returns a result for anticipated and actual binary classes between -1 and +1.

5.3. Average F1-Score

The F1 score was used to gauge the effectiveness of the classification model in a study that required forecasting binary classifications. In classification difficulties, the F1-score, which assesses the equilibrium between precision and recall, is frequently used. The number of accurate predictions the model made was taken into account while calculating the average F1-score, which was used to gauge the model's performance. This method of using the average F1-score as a criterion for estimating the performance of the categorization model was employed in the study.
Since the dataset for vehicle-pedestrian crashes is so unbalanced, additional evaluation criteria beyond general correctness must be taken into account. In this study, additional accepted criteria were used to assess how well BNN architectures performed. However, the aforementioned three parameters served as the foundation for the final evaluations and comparisons. Classification accuracy, sensitivity, specificity, negative predictive value, false predictive value, true positive rate, and false positive rate were included as additional criteria. The followings are the equations that are used to compute specificity, sensitivity, and MCC.

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

Where TN (True Negative) is the number of cases that are truly negative and correctly classified as negative by the test, and FP (False Positive) is the number of cases that are truly negative but classified as positive by the test.

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \tag{5}$$

Where True Positives (TP) are the number of correctly predicted positive cases and False Negatives (FN) are the numbers of wrongly predicted negative cases.

$$\text{MCC} = \frac{(\text{TP} \times \text{TN} - \text{FP} \times \text{FN})}{\sqrt{(\text{TP} + \text{FP}) \times (\text{TP} + \text{FN}) \times (\text{TN} + \text{FP}) \times (\text{TN} + \text{FN})}} \tag{6}$$

Where TP represents True Positive, TN represents True Negative, FP represents False Positive, and FN is for False Negative.

6. Results and Discussion

6.1. Determination of Significant variables

Using the sophisticated chi-square (χ^2) approach, this study refined unrealistic inputs for a BNN vehicle-pedestrian collision model. A statistical test called chi-square is used to assess whether there is a

substantial discrepancy between the noted and predicted regularities of a categorical variable. The chi-square value is computed to establish the significance score (weight) of each variable. It's worth noting that the chi-square test can only be used when the sample size is large and when the expected frequencies are greater than 5 in all cells. A variable that contains more information for categorization is given a higher chi-square value. This criterion is capable of fairly taking into account both categorical and continuous inputs to calculate and order the inputs. On 8 variables, the authors used the chi-square algorithm. These variables include junction control, weather condition, hit and run, light conditions during the accident, district at which the accident occurred, rain, year, and movement of the vehicle. The significant variable is ordered in the following Figure 4.
Figure 4 illustrates that vehicle movement in any district has a significant influence on the severity of vehicle-pedestrian collisions. The "one-vehicle movement" causes most of the crashes in Hong Kong. This might be because of driver inattention or distraction. Drivers may not be paying attention to the road and driving at a high-speed while may not see pedestrians crossing or walking on the side of the road. The natural light condition has the least impact on the extremity of vehicle-pedestrian collisions. According to the present research, the risk of fatal and severe vehicle-pedestrian accidents at intersections decreases when it is daylight outside. This is likely since visibility is better during the day. However, it is interesting to note that the risk of serious injury to pedestrians is also minimized at night, when visibility may be lower. This may be because the region has streetlights and good illumination, which aids in maintaining visibility even in dimly lit outdoor areas.

6.2. Performance Assessment of BNN and NN Model

Initially, the BNN and NN models were developed, and their assessment criteria are discussed in Table 2 when the repeat train test value was set as 10 and the set size for training of models was considered 70%. The performance of both BNN and NN are contrasted and the authors of this research put an effort to gauge the performance of these models against other machine learning techniques like random forest (RF) and K-Nearest Neighbors (KNN). The table shows the AUC, classification accuracy (CA), and the precision value obtained during modeling. The outcome of this study revealed that BNN techniques outperformed NN, RF, and KNN.

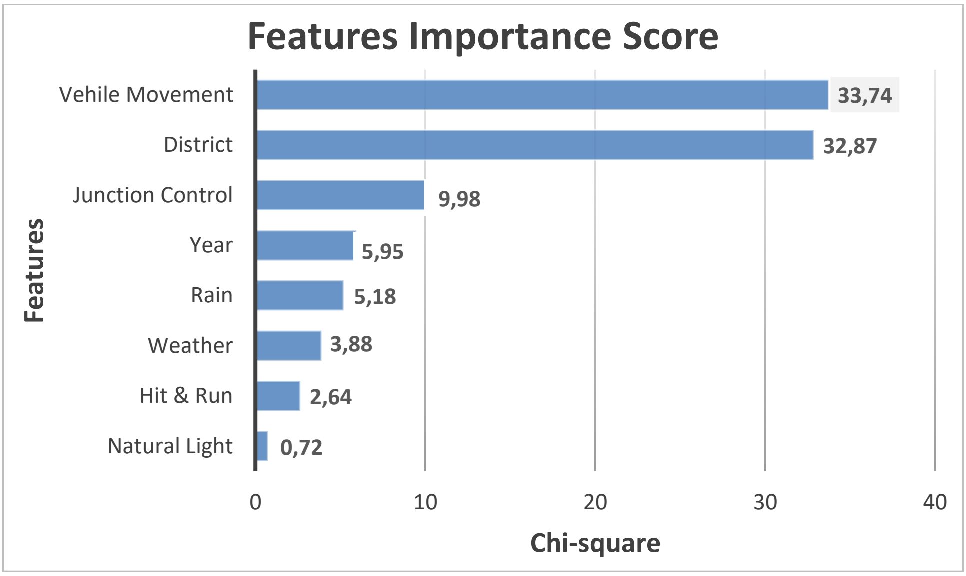

Fig. 4. Sensitivity analysis chart

Table 2. Comparisons of the BNN model with other ML models

Models	AUC	CA	Precision
BNN	0.789	0.765	0.698
NN	0.571	0.745	0.655
RF	0.577	0.753	0.659
KNN	0.572	0.754	0.584

However, these results were optimized using the repeat train test value of 100. By incorporating more data for training, the model has the opportunity to learn more about the underlying patterns in the data and the model is less likely to overfit on the training data, as it is exposed to the same data multiple times which ultimately leads to better performance. Table 3 presents the improved results of AUC, CA, and precision which can be concluded as BNN outperformed as compared to other techniques. However standard NN showed the least accurate prediction performance as compared to other ML techniques.

Table 3. Comparisons of the BNN model with other ML models with improved results

Models	AUC	CA	Precision
BNN	0.897	0.764	0.682
NN	0.643	0.746	0.655
RF	0.670	0.753	0.655
KNN	0.662	0.764	0.584

6.3. Quantification of Models

The BNN and NN models are compared with each other based on assessment criteria that include area under the ROC curve, classification accuracy, specificity, precision, sensitivity, F-1 score, true positive rate, negative predictive rate, positive predictive rate, false positive rate, and Mathew's correlation coefficient. The results are concluded in Table 4 which can be summarized as the BNN model outperformed NN predictions. The AUC, MCC, and F1-score values at all the severity levels of BNN models are greater than the standard NN model. It can also be noticed from the table that the Bayesian model forecast the pedestrian crash data of slight severity with a high degree of accuracy and a similar pattern is observed in the Neural Network model. As from the Table 1 the total number of fatal accidents is 200 which is far less than an accident with the severity of slight or serious injuries hence a smaller number of data points leads to the inaccuracy of models. The Bayesian method is a statistical technique that is notable for two key characteristics. First, it provides predictions in the form of probabilities for each class, rather than simply giving a definitive class label. This allows for a more nuanced understanding of the model's predictions. Second, it also provides an estimate of the level of uncertainty of the prediction by providing the standard deviation of the posterior prediction. This indicates how much confidence one can have in the prediction and how much it can vary.

Table 4. Performance of the BNN model and NN model

Models	**BNN**			**NN**		
Severity Level	**Fatal**	**Serious**	**Slight**	**Fatal**	**Serious**	**Slight**
AUC	0.888	0.892	0.897	0.601	0.626	0.643
CA	0.977	0.786	0.766	0.974	0.763	0.740
Specificity	0.999	0.999	0.023	0.995	0.947	0.101
Precision	0.682	0.681	0.682	0.674	0.675	0.676
Sensitivity	0.000	0.003	0.995	0.007	0.086	0.937
True Positive Rate	0.000	0.003	0.995	0.007	0.086	0.937
False Positive Rate	0.001	0.001	0.977	0.005	0.053	0.899
Positive Predictive Value	0.000	0.360	0.767	0.032	0.307	0.772
Negative Predictive Value	0.978	0.786	0.584	0.978	0.792	0.333
F1-Score	0.693	0.699	0.706	0.666	0.667	0.674
MCC	0.540	0.598	0.786	0.320	0.405	0.457

The graph in Figure 5(a,b,c) of the F1 score provides a visual representation of the trade-off between precision and recall. For example, a model that has high precision, but low recall would result in a low F1 score because the model is not correctly identifying all the positive instances. On the other hand, a model that has a high recall, but low precision would result in a low F1 score because it is generating many false positive predictions. The goal is to have a model with a high F1 score, indicating that it has both a high precision value and a high recall value.
A graph presented in Figure 5(d,e,f) of PPV and NPV is often used to analyze the performance of a model across different classification thresholds, which can be adjusted to trade-off between precision and recall. A perfect model will have a PPV and NPV of 1 and will therefore be represented as a point at the top-right corner of the graph, where the x and y axes meet. Conversely, a random model will have a PPV and NPV of 0.5 and will be represented as a point in the center of the graph. The shape of the curve in the PPV vs. NPV graph can also provide information about the model's performance.
The graph in Figure 5(g,h,i) provides a visual representation of the trade-off between specificity and sensitivity and enables us to determine the optimal threshold for a binary classification model. A threshold is a boundary that separates positive and negative cases based on the model's predicted probabilities or scores. The optimal threshold is the point on the specificity-sensitivity graph that maximizes the overall accuracy of the model. It is the balance point between detecting all positive cases (high sensitivity) and avoiding false positive cases (high specificity). A threshold that is set too high will result in a large number of false negatives since sensitivity will drop and specificity will rise. When the threshold is set too low, the sensitivity and specificity both rise, producing a lot of false positives.
The TPR and FPR graphs shown in Figure 5(j,k,l) are useful for selecting a suitable threshold for the binary classifier, balancing the trade-off between correctly classifying positive samples and avoiding false positive classifications. They also provide the procedure to compare the performance of multiple classifiers and select the one that gives the best balance between TPR and FPR. The top left corner of the ROC graph would include a point representing a perfect classifier with a TPR of 1.0 and an FPR of 0.0.

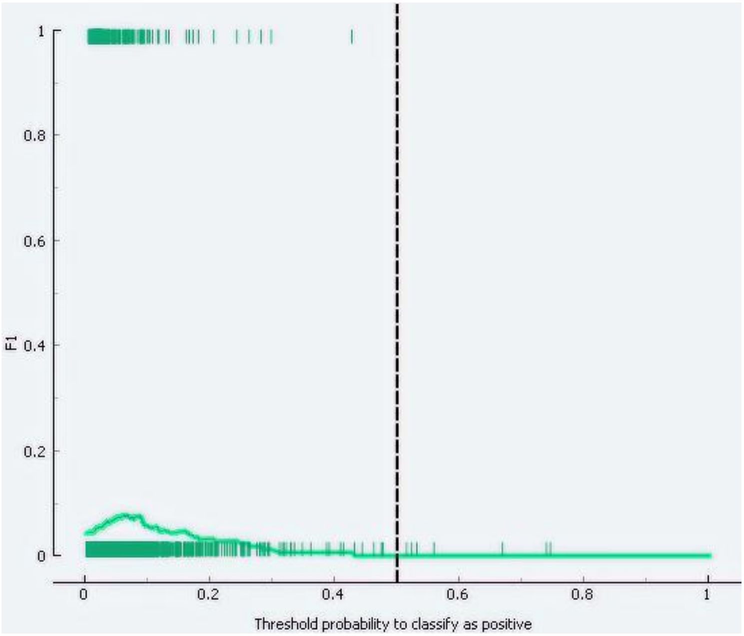

Fig. 5. (a) F1-score of fatal severity level

The classification accuracy graph as shown in Figure 5(m,n,o) can be used to determine the best threshold value for classifying the cases based on their predicted probabilities. The graph provides a clear visual representation of the trade-off between the sensitivity and specificity of the classifier, allowing practitioners to make informed decisions about the classifier's performance. It is used to assess how well a classifier can predict the class labels of a given dataset.

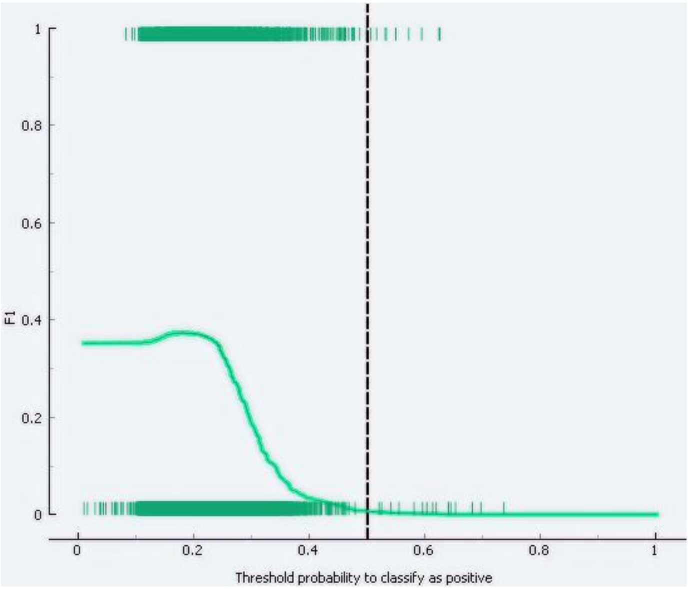

Fig. 5. (b) F1-score of serious severity level

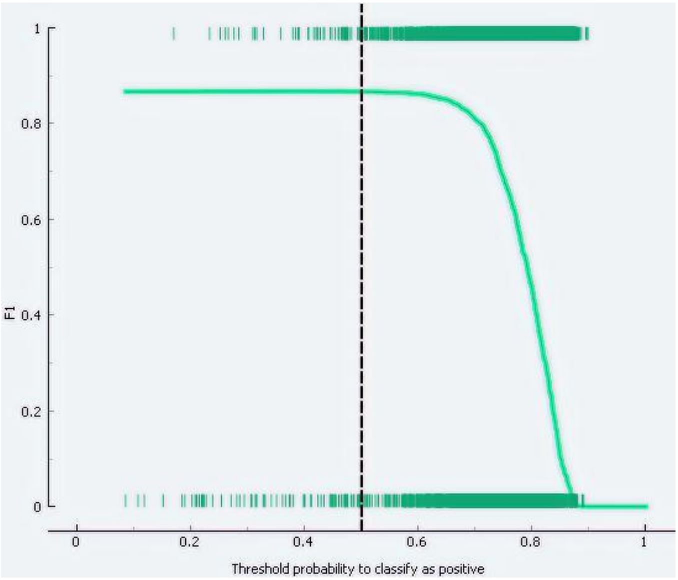

Fig. 5. (c) F1-score of slight severity level

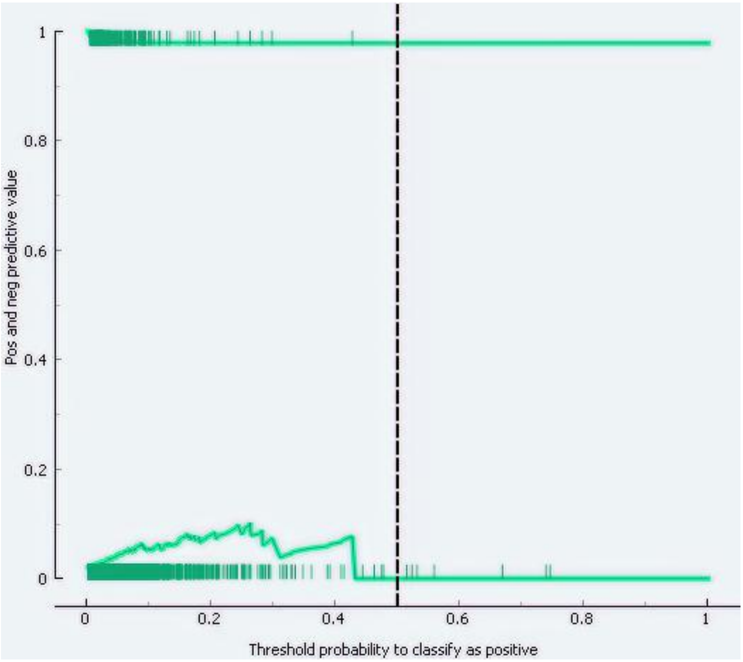

Fig. 5. (d) Positive and Negative predictive value for fatal severity

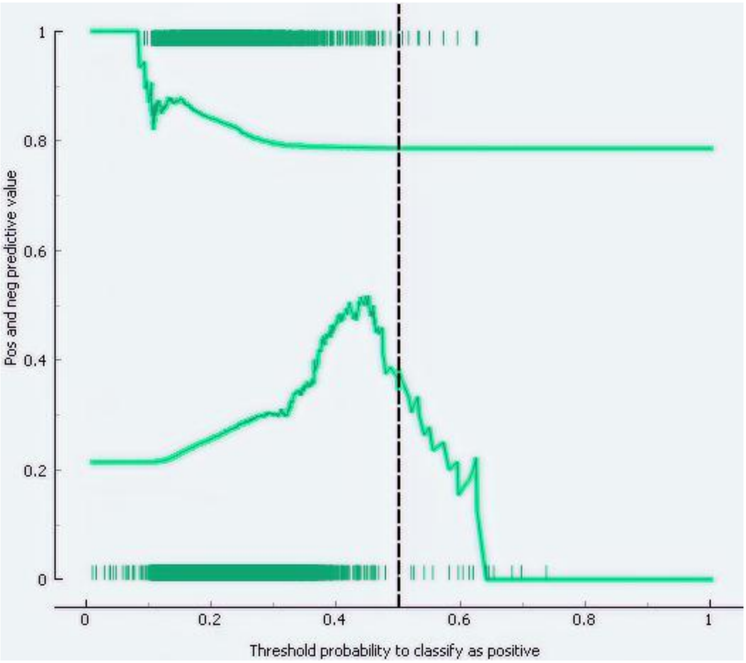

Fig. 5. (e) Positive and Negative predictive value for serious severity

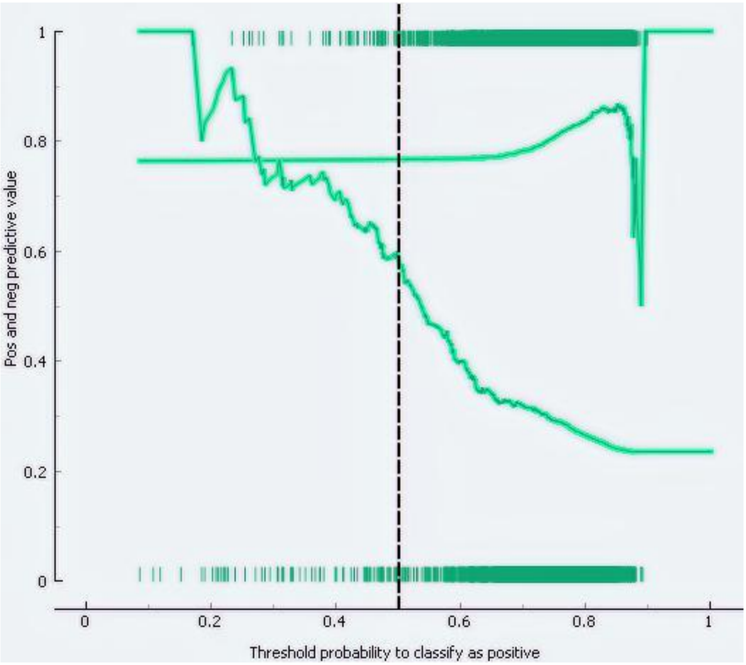

Fig. 5. (f) Positive and Negative predictive value for slight severity

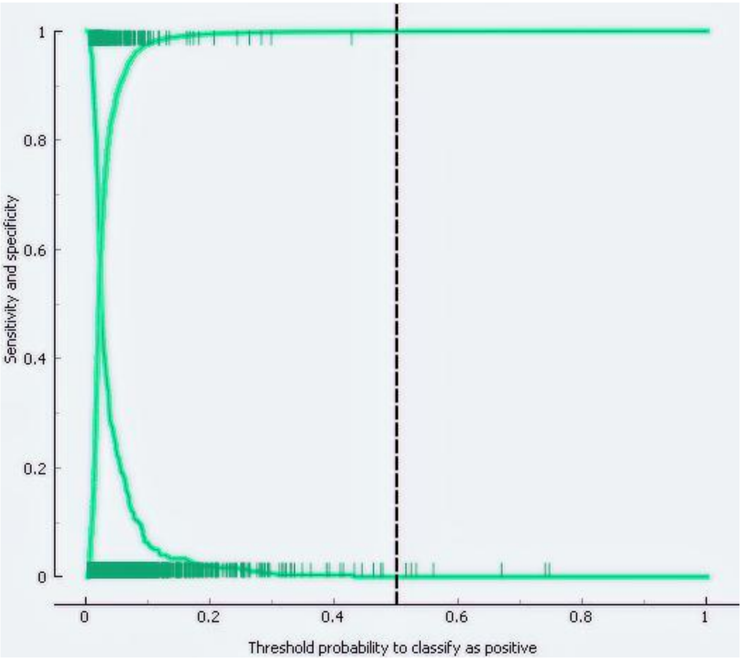

Fig. 5. (g) Sensitivity and specificity for fatal severity

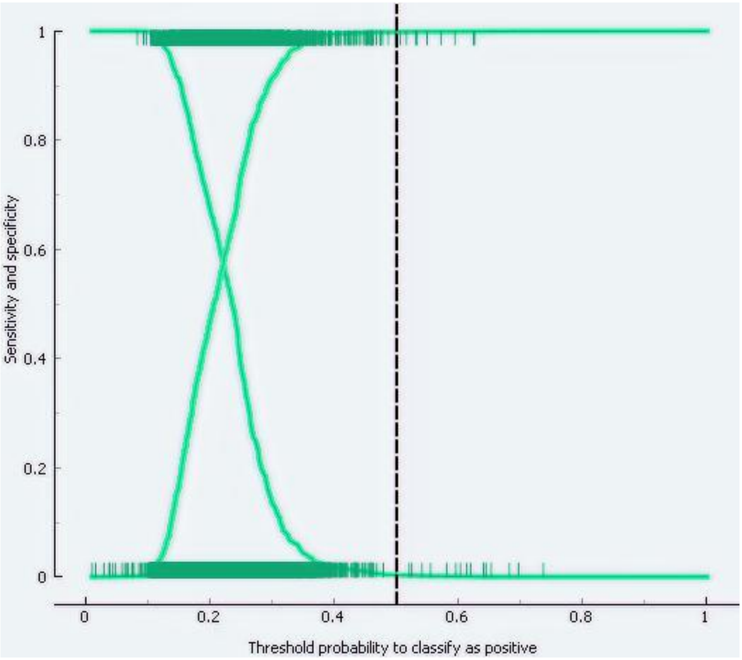

Fig. 5. (h) Sensitivity and specificity for serious severity

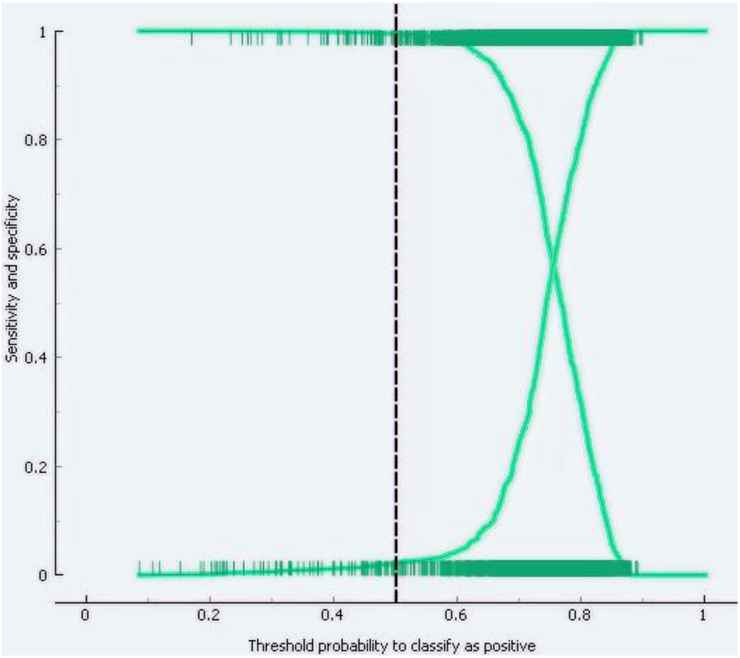

Fig. 5. (i) Sensitivity and specificity for slight severity

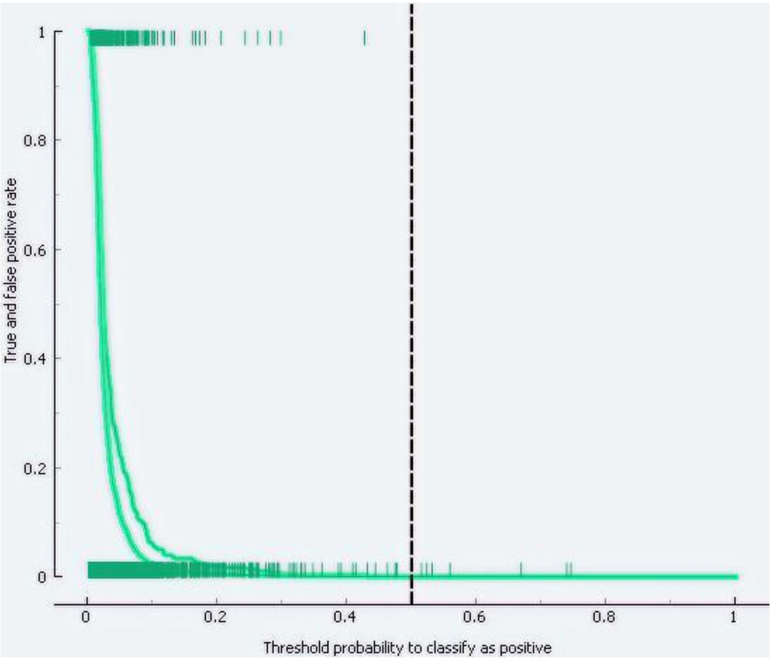

Fig. 5. (j) True and False positive rate for fatal severity

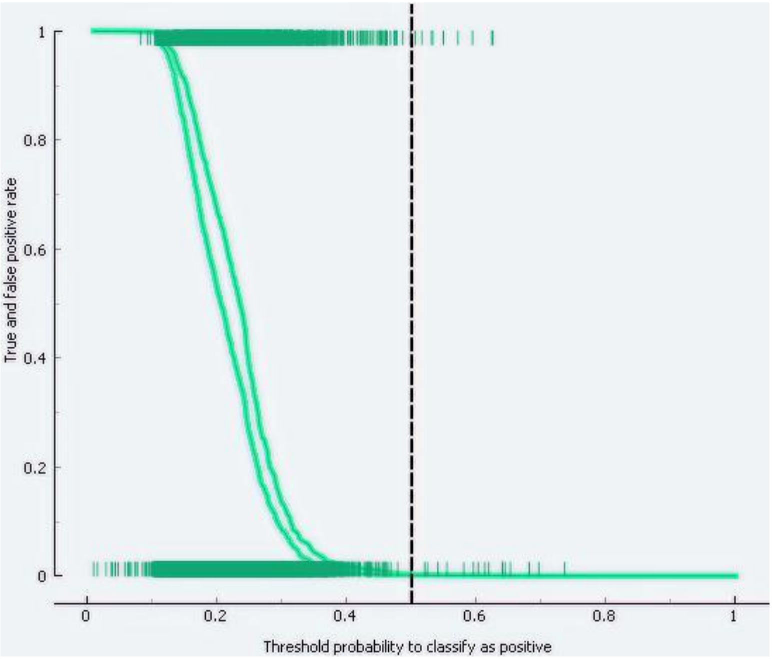

Fig. 5. (k) True and False positive rate for serious severity

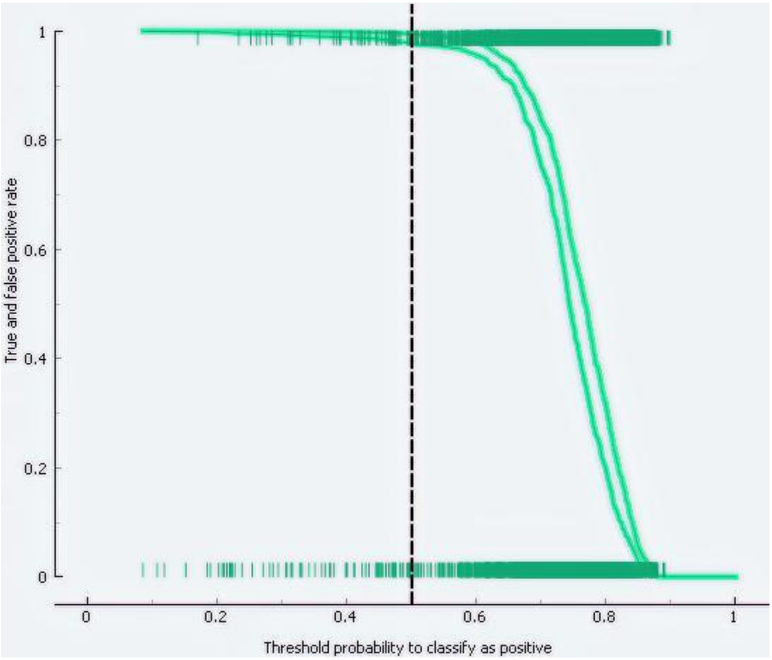

Fig. 5. (l) True and False positive rate for slight severity

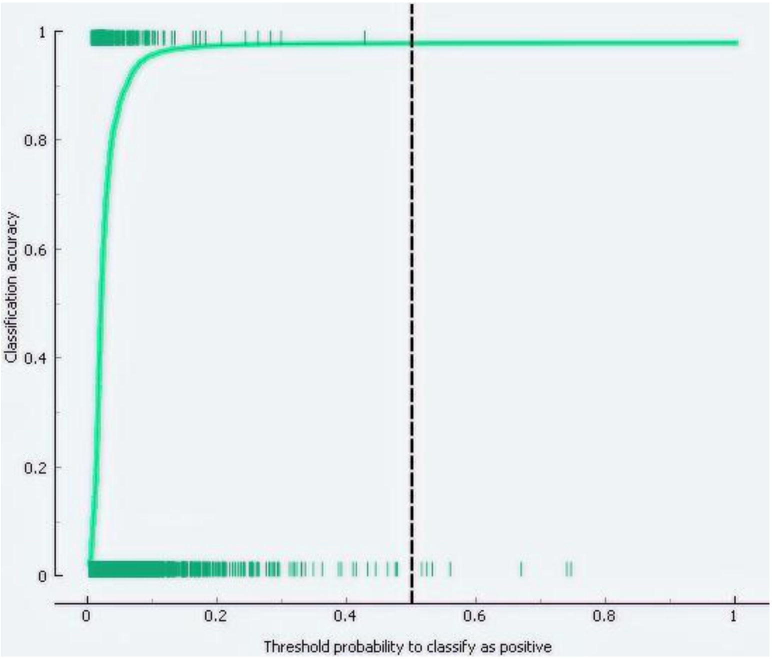

Fig. 5. (m) Class accuracy for fatal severity

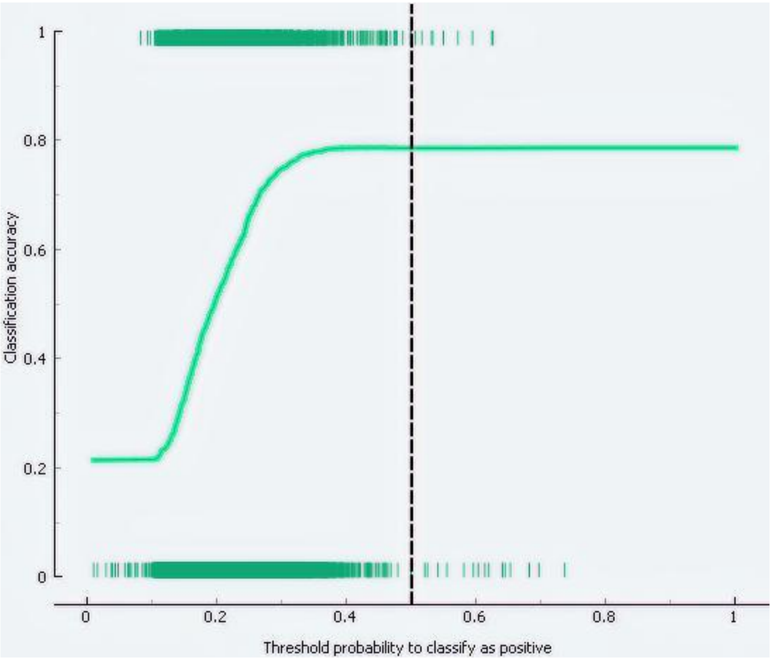

Fig. 5. (n) Class accuracy for serious severity

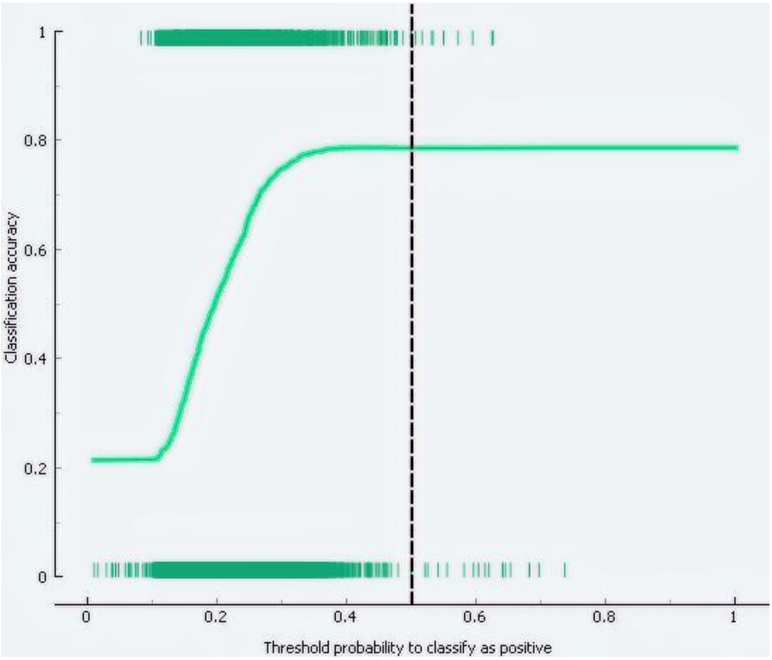

Fig. 5. (o) Class accuracy for slight severity.

6.4. Features Significance According to BNN Model Predictions

After performing the Bayesian analysis on the Hong Kong vehicle-pedestrian crashes data set the significant features that affect the severity of the accidents are summarized in Figure 6. The graph shows the correlation between numerous variables, including traffic regulation, environmental conditions, and geography against the extremity of vehicle-pedestrian collisions. These studies also show that the presence of traffic signals and the lack of control tend to enhance the possibility of serious vehicle-pedestrian crashes at crossings (Liu et al., 2019, Koepsell et al., 2002, Sze and Wong, 2007). This may be because drivers and pedestrians are more cautious when traffic signals are not present. On contrary, the possibility of a severe injury may increase if no control is available at a junction. Additionally, it has been discovered that light rain maximizes the probability of fatal and severe vehicle-pedestrian collisions because it impairs visibility and reaction time. The severity of crashes is not, however, significantly predicted by heavy rain (Tay et al., 2011, Amoh-Gyimah et al., 2017). The graph also indicates that the risk of deadly and serious vehicle-pedestrian crashes is reduced during daylight hours and when visibility is good due to the presence of streetlights (Park et al., 2012, Aziz et al., 2013, Kim et al., 2010, Amoh-Gyimah et al., 2017.

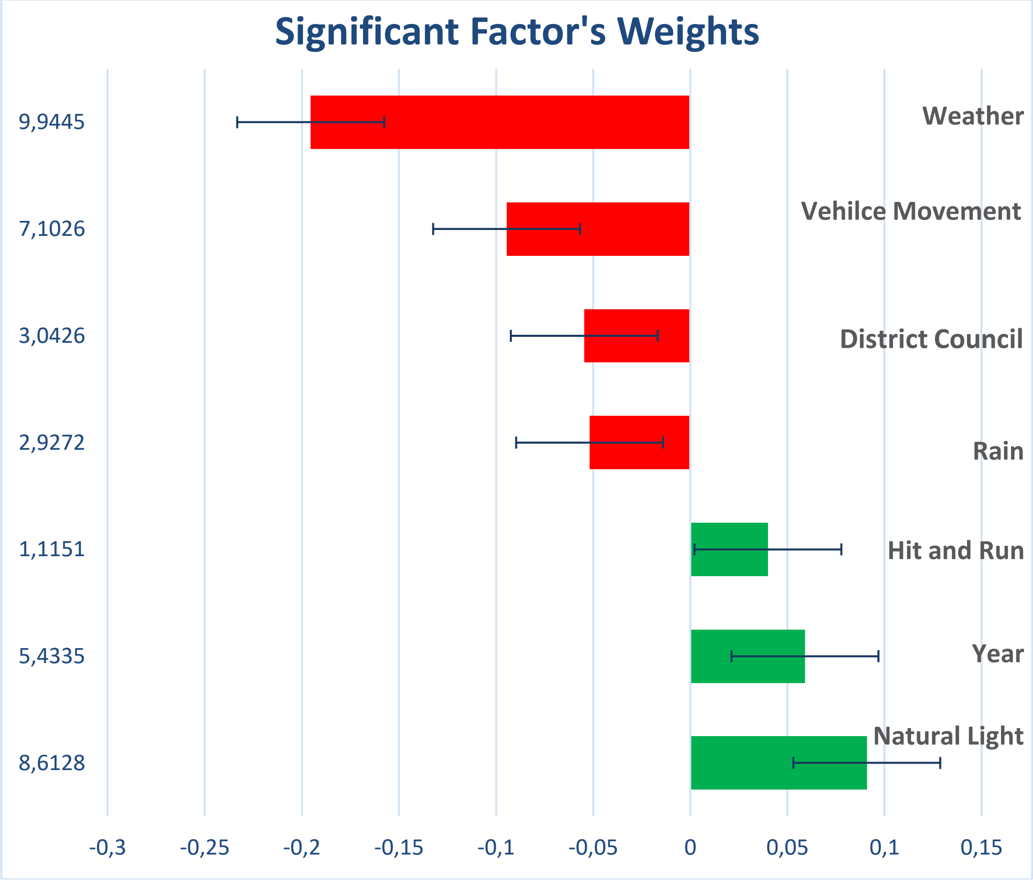

Figure 6. Weights of significant factors

7. Conclusion

In this study, the BNN model was utilized using data from the Hong Kong transportation department database to generate accurate predictions of pedestrian fatalities and injuries based on numerous factors such as weather conditions, time of the crash, location, road conditions, and crash attributes. Different BNN model structures were created and tested for each class of crash severity to evaluate their performance and to understand the level of uncertainty in their predictions. The study's results are summarized below.

- The BNN model shows the best performance against NN, SVM, and KNN models with the area under the ROC curve and precision value of 0.897 and 0.682 respectively.
- The use of weather characteristics, time, and occasion factors as input data significantly improves the performance of the model, making them crucial elements in the model.
- Vehicle movement and junction control were found to be the most crucial individual factors in forecasting vehicle-pedestrian crashes.
- Improved road characteristics and making intersections more signalized in Hong Kong reduces the risk of vehicle crashes.

To minimize the number of accidents involving vehicles and pedestrians, several measures are suggested. One of these is to ensure good visibility, particularly during dawn and dusk, by providing adequate lighting. Additionally, it is recommended that both drivers and pedestrians take extra care and pay more attention when the weather is poor. Another solution is to conduct campaigns to educate both pedestrians and drivers about the importance of following traffic laws even when there are traffic signals present. Furthermore, it is important to focus on areas where there is a higher likelihood of fatal and severe accidents involving vehicles and pedestrians. However, it is noteworthy that the findings of this research are based on the data that is available, and more factors that contribute to accidents could be studied in future research.

References

[1] Abdel-Aty, M., Keller, J. (2005). Exploring the overall and specific crash severity levels at signalized intersections. Accident Analysis, 37, 417-425. https://doi.org/10.1016/j.aap.2004.11.002.

[2] Aghaabbasi, M., Shekari, Z. A., Shah, M. Z., Olakunle, O., Armaghani, D. J., Moeinaddini, M. (2020). Predicting the use frequency of ride-sourcing by off-campus university students through random forest and Bayesian network techniques. Transportation Research Part A: Policy, 136, 262-281. https://doi.org/10.53623/csue.v1i1.28.

[3] Aghaabbasi, M., Ali, M., Jasiński, M., Leonowicz, Z., Novák, T. (2023). On Hyperparameter Optimization of Machine Learning Methods Using a Bayesian Optimization Algorithm to Predict Work Travel Mode Choice. IEEE Access, 11, 19762-19774. https://doi.org/10.1109/ACCESS.2023.3247448.

[4] Al-Ghamdi, A. S. (2002). Using logistic regression to estimate the influence of accident factors on accident severity. Accident Analysis, Prevention, 34, 729-741. https://doi.org/10.1016/S0001-4575(01)00073-2.

[5] Al Mamlook, R. E., Abdulhameed, T. Z., Hasan, R., Al-Shaikhli, H. I., Mohammed, I., Tabatabai, S. (2020). Utilizing machine learning models to predict the car crash injury severity among elderly drivers. 2020 IEEE international conference on electro information technology (EIT). IEEE, 105-111. https://doi.org/10.1109/EIT48999.2020.9208259.

[6] Ali, M., Abbas, S., Salah, B., Akhter, J., Saleem, W., Haruna, S., Room, S., Abdulkadir, I. (2021). Investigating Optimal Confinement Behaviour of Low-Strength Concrete through Quantitative and Analytical Approaches. Materials, 14, 4675. https://doi.org/10.3390/ma14164675.

[7] Ali, M., Hin Lai, S. (2023). Artificial intelligent techniques for prediction of rock strength and deformation properties – A review. Structures, 55, 1542-1555. https://doi.org/10.1016/j.istruc.2023.06.131.

[8] Ali, Y., Haque, M. M., Mannering, F. (2023). A Bayesian generalised extreme value model to estimate real-time pedestrian crash risks at signalised intersections using artificial intelligence-based video analytics. Analytic methods in accident research, 38, 100264. https://doi.org/10.1016/j.amar.2022.100264.

[9] Ali, M., Dharmowijoyo, D. B. E., De Azevedo, A. R. G., Fediuk, R., Ahmad, H., Salah, B. (2021). Time-Use and Spatio-Temporal Variables Influence on Physical Activity Intensity, Physical and Social Health of Travelers. Sustainability, 13, 12226. https://doi.org/10.3390/su132112226.
[10] Amoh-Gyimah, R., Aidoo, E. N., Akaateba, M. A., Appiah, S. K. (2017). The effect of natural and built environmental characteristics on pedestrian-vehicle crash severity in Ghana. International journal of injury control safety promotion, 24, 459-468. https://doi.org/10.1080/17457300.2016.1232274.
[11] Arifeen, S. U., Amin, M. N., Ahmad, W., Althoey, F., Ali, M., Alotaibi, B. S., Abuhussain, M. A. (2023). A comparative study of prediction models for alkali-activated materials to promote quick and economical adaptability in the building sector. Construction and Building Materials, 407, 133485. https://doi.org/10.1016/j.conbuildmat.2023.133485.
[12] Astarita, V., Haghshenas, S. S., Guido, G., Vitale, A. (2023). Developing new hybrid grey wolf optimization-based artificial neural network for predicting road crash severity. Transportation Engineering, 12, 100164. https://doi.org/10.1016/j.treng.2023.100164.
[13] Aziz, H. A., Ukkusuri, S. V., Hasan, S. (2013). Exploring the determinants of pedestrian–vehicle crash severity in New York City. Accident Analysis, Prevention, 50, 1298-1309. https://doi.org/10.1016/j.aap.2012.09.034.
[14] 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. https://doi.org/10.1016/j.amar.2016.07.002.
[15] Chen, Y., Aghaabbasi, M., Ali, M., Anciferov, S., Sabitov, L., Chebotarev, S., Nabiullina, K., Sychev, E., Fediuk, R., Zainol, R. (2021). Hybrid Bayesian Network Models to Investigate the Impact of Built Environment Experience before Adulthood on Students' Tolerable Travel Time to Campus: Towards Sustainable Commute Behavior. Sustainability, 14, 325. https://doi.org/10.3390/su14010325.
[16] Chen, Z. F. (2019). Modeling pedestrian injury severity in pedestrian-vehicle crashes in rural and urban areas: mixed logit model approach. Transportation research record 2673, 1023-1034. https://doi.org/10.1177/0361198119842825.
[17] Dai, D. (2012). Identifying clusters and risk factors of injuries in pedestrian–vehicle crashes in a GIS environment. Journal of Transport Geography, 24, 206-214. https://doi.org/10.1016/j.jtrangeo.2012.02.005.
[18] Das, S., Le, M., Dai, B. (2020). Application of machine learning tools in classifying pedestrian crash types: A case study. Transportation Safety and Environment, 2, 106-119. https://doi.org/10.1093/tse/tdaa010.
[19] De Lavalette, B. C., Tijus, C., Poitrenaud, S., Leproux, C., Bergeron, J., Thouez, J.-P. (2009). Pedestrian crossing decision-making: A situational and behavioral approach. Safety science, 47, 1248-1253. https://doi.org/10.1016/j.ssci.2009.03.016.
[20] Ding, C., Chen, P., Jiao, J. (2018). Non-linear effects of the built environment on automobile-involved pedestrian crash frequency: A machine learning approach. Accident Analysis, Prevention, 112, 116-126. https://doi.org/10.1016/j.aap.2017.12.026.
[21] Faraz, M. I., Arifeen, S. U., Amin, M. N., Nafees, A., Althoey, F., Niaz, A. (2023). A comprehensive GEP and MEP analysis of a cement-based concrete containing metakaolin. Elsevier, 937-948. https://doi.org/10.1016/j.istruc.2023.04.050.
[22] Gårder, P. E. (2004). The impact of speed and other variables on pedestrian safety in Maine. Accident Analysis, Prevention, 36, 533-542. https://doi.org/10.1016/S0001-4575(03)00059-9.
[23] 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, 926. https://doi.org/10.3390/su13020926.
[24] Hafeez, F., Sheikh, U. U., Al-Shammari, S., Hamid, M., Khakwani, A. B. K., Arfeen, Z. A. (2023). Comparative analysis of influencing factors on pedestrian road accidents. Bulletin of Electrical Engineering and Informatics, 12,

257-267. https://doi.org/10.11591/eei.v12i1.4312.

[25] Hilakivi, I., Veilahti, J., Asplund, P., Sinivuo, J., Laitinen, L., Koskenvuo, K. (1989). A sixteen-factor personality test for predicting automobile driving accidents of young drivers. Accident Analysis, Prevention, 21, 413-418. https://doi.org/10.1016/0001-4575(89)90001-8.

[26] James, J. L., Kim, K. E. (1996). Restraint use by children involved in crashes in Hawaii, 1986–1991. Transportation research record, 1560, 8-12. https://doi.org/10.1177/0361198196156000102.

[27] Jian, L., Lizhong, Y., Daoliang, Z. (2005). Simulation of bi-direction pedestrian movement in corridor. Physica A: Statistical Mechanics and its Applications, 354, 619-628. https://doi.org/10.1016/j.physa.2005.03.007.

[28] Khuzan, T. S., Al-Jumaili, M. A. (2023). A review of studying the relationship of rural road accidents with geometric design. AIP Publishing. https://doi.org/10.1063/5.0150006.

[29] Kim, D.-G., Washington, S., Oh, J. (2006). Modeling crash types: New insights into the effects of covariates on crashes at rural intersections. Journal of Transportation Engineering, 132, 282-292. https://doi.org/10.1061/(ASCE)0733-947X(2006)132:4(282).

[30] Kim, J.-K., Ulfarsson, G. F., Shankar, V. N., Mannering, F. L. (2010). A note on modeling pedestrian-injury severity in motor-vehicle crashes with the mixed logit model. Accident Analysis, Prevention, 42, 1751-1758. https://doi.org/10.1016/j.aap.2010.04.016.

[31] Koepsell, T., Mccloskey, L., Wolf, M., Moudon, A. V., Buchner, D., Kraus, J., Patterson, M. (2002). Crosswalk markings and the risk of pedestrian–motor vehicle collisions in older pedestrians. Jama, 288, 2136-2143. https://doi.org/doi:10.1001/jama.288.17.2136.

[32] Lee, S., Kim, S., Kim, J., Kim, D., Lee, D., Im, G., Yuk, H., Heo, T.-Y. (2023). Multiclass Classification by Various Machine Learning Algorithms and Interpretation of the Risk Factors of Pedestrian Accidents Using Explainable AI. Mathematical Problems in Engineering, 2023 https://doi.org/10.1155/2023/1956865.

[33] Li, D., Ranjitkar, P., Zhao, Y., Yi, H., Rashidi, S. (2017). Analyzing pedestrian crash injury severity under different weather conditions. Traffic injury prevention, 18, 427-430 https://doi.org/10.1080/15389588.2016.1207762.

[34] Li, Y., Ma, D., Zhu, M., Zeng, Z., Wang, Y. (2018). Identification of significant factors in fatal-injury highway crashes using genetic algorithm and neural network. Accident Analysis, Prevention, 111, 354-363. https://doi.org/10.1016/j.aap.2017.11.028.

[35] Liang, M., Li, S. (2022). Enhancing BNN structure learning of hybrid modeling strategy for free formulated mechanism complex systems. Journal of Process Control, 120, 44-67. https://doi.org/10.1016/j.jprocont.2022.10.006.

[36] Liu, K., Yu, Q., Yuan, Z., Yang, Z., Shu, Y. (2021). A systematic analysis for maritime accidents causation in Chinese coastal waters using machine learning approaches. Ocean, Coastal Management, 213, 105859. https://doi.org/10.1016/j.ocecoaman.2021.105859.

[37] Liu, T., Liu, Y., Liu, J., Wang, L., Xu, L., Qiu, G., Gao, H. (2020). A Bayesian learning based scheme for online dynamic security assessment and preventive control. IEEE Transactions on Power Systems, 35, 4088-4099. https://doi.org/10.1109/TPWRS.2020.2983477.

[38] Lu, W., Liu, J., Fu, X., Yang, J., Jones, S. (2022). Integrating machine learning into path analysis for quantifying behavioral pathways in bicycle-motor vehicle crashes. Accident Analysis, Prevention, 168, 106622. https://doi.org/10.1016/j.aap.2022.106622.

[39] Luke, R. (2023). Current and future trends in driver behaviour and traffic safety scholarship: an African research agenda. International journal of environmental research and public health, 20, 4290. https://doi.org/10.3390/ijerph20054290.

[40] Ma, T., Aghaabbasi, M., Ali, M., Zainol, R., Jan, A., Mohamed, A. M., Mohamed, A. (2022). Nonlinear Relationships between Vehicle Ownership and Household Travel

Characteristics and Built Environment Attributes in the US Using the XGBT Algorithm. Sustainability, 14, 3395. https://doi.org/10.3390/su14063395.

[41] Mafi, S., Abdelrazig, Y., Doczy, R. (2018). Machine learning methods to analyze injury severity of drivers from different age and gender groups. Transportation research record, 2672, 171-183. https://doi.org/10.1177/0361198118794292.

[42] Mannering, F. L., Grodsky, L. L. (1995). Statistical analysis of motorcyclists' perceived accident risk. Accident Analysis, Prevention, 27, 21-31. https://doi.org/10.1016/0001-4575(94)00041-J.

[43] Marzban, C., Witt, A. (2001). A Bayesian neural network for severe-hail size prediction. Weather and Forecasting, 16, 600-610. https://doi.org/10.1175/1520-0434(2001)016%3C0600:ABNNFS%3E2.0.CO;2.

[44] Marzoug, R., Lakouari, N., Ez-Zahraouy, H., Téllez, B. C., Téllez, M. C., Villalobos, L. C. (2022). Modeling and simulation of car accidents at a signalized intersection using cellular automata. Physica A: Statistical Mechanics and its Applications, 589, 126599. https://doi.org/10.1016/j.physa.2021.126599.

[45] Maze, T. H., Agarwal, M., Burchett, G. (2006). Whether weather matters to traffic demand, traffic safety, and traffic operations and flow. Transportation research record, 1948, 170-176. https://doi.org/10.1177/0361198106194800119.

[46] Mercier, C. R., Shelley, M. C., Rimkus, J. B., Mercier, J. M. (1997). Age and gender as predictors of injury severity in head-on highway vehicular collisions. Transportation Research Record, 1581, 37-46. https://doi.org/10.3141/1581-05.

[47] Mohamed, M. G., Saunier, N., Miranda-Moreno, L. F., Ukkusuri, S. V. (2013). A clustering regression approach: A comprehensive injury severity analysis of pedestrian–vehicle crashes in New York, US and Montreal, Canada. Safety science, 54, 27-37. https://doi.org/10.1016/j.ssci.2012.11.001.

[48] Mondal, A. R., Bhuiyan, M. A. E., Yang, F. (2020). Advancement of weather-related crash prediction model using nonparametric machine learning algorithms. SN Applied Sciences, 2, 1-11. https://doi.org/10.1007/s42452-020-03196-x.

[49] Nayeem, M. A., Hasan, A. S., Jalayer, M. (2023). Investigation of Young Pedestrian Crashes in School Districts of New Jersey Using Machine Learning Models. 250-264. https://doi.org/10.1061/9780784484883.022.

[50] Olowosegun, A., Babajide, N., Akintola, A., Fountas, G., Fonzone, A. (2022). Analysis of pedestrian accident injury-severities at road junctions and crossings using an advanced random parameter modelling framework: The case of Scotland. Accident Analysis, Prevention, 169, 106610. https://doi.org/10.1016/j.aap.2022.106610.

[51] Pucher, J., Dijkstra, L. (2003). Promoting safe walking and cycling to improve public health: lessons from the Netherlands and Germany. American journal of public health, 93, 1509-1516. https://doi.org/10.2105/AJPH.93.9.1509.

[52] Qian, Y., Aghaabbasi, M., Ali, M., Alqurashi, M., Salah, B., Zainol, R., Moeinaddini, M., Hussein, E. E. (2021). Classification of Imbalanced Travel Mode Choice to Work Data Using Adjustable SVM Model. Applied Sciences, 11, 11916. https://doi.org/10.3390/app112411916.

[53] Quddus, M. A., Noland, R. B., Chin, H. C. (2002). An analysis of motorcycle injury and vehicle damage severity using ordered probit models. Journal of Safety research, 33, 445-462. https://doi.org/10.1016/S0022-4375(02)00051-8.

[54] Rahimi, A., Azimi, G., Asgari, H., Jin, X, (2020). Injury severity of pedestrian and bicyclist crashes involving large trucks. International Conference on Transportation and Development. American Society of Civil Engineers Reston, VA, 110-122. https://doi.org/https://ascelibrary.org/doi/abs/10.1061/9780784483152.010.

[55] Saha, D., Dumbaugh, E. (2021). Use of a model-based gradient boosting framework to assess spatial and non-linear effects of variables on pedestrian crash frequency at macro-level. Journal of Transportation Safety, Security, 1-32. https://doi.org/10.1080/19439962.2021.1958036.

[56] 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, 1535-1547. https://doi.org/10.1007/s00521-022-07769-2.
[57] Shankar, V., Mannering, F. (1996). An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity. Journal of safety research, 27, 183-194. https://doi.org/10.1016/0022-4375(96)00010-2.
[58] Sze, N.-N., Wong, S. (2007). Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes. Accident Analysis, Prevention, 39, 1267-1278. https://doi.org/10.1016/j.aap.2007.03.017.
[59] Tao, W., Aghaabbasi, M., Ali, M., Almaliki, A. H., Zainol, R., Almaliki, A. A., Hussein, E. E. (2022). An advanced machine learning approach to predicting pedestrian fatality caused by road crashes: A step toward sustainable pedestrian safety. Sustainability, 14, 2436. https://doi.org/10.3390/su14042436.
[60] Tay, R., Choi, J., Kattan, L., Khan, A. (2011). A multinomial logit model of pedestrian–vehicle crash severity. International journal of sustainable transportation, 5, 233-249. https://doi.org/10.1080/15568318.2010.497547.
[61] Tang, P., Aghaabbasi, M., Ali, M., Jan, A., Mohamed, A. M., Mohamed, A. (2022). How Sustainable Is People Travel to Reach Public Transit Stations to Go to Work? A Machine Learning Approach to Reveal Complex Relationships. Sustainability, 14, 3989. https://doi.org/10.3390/su14073989.
[62] Theofilatos, A., Yannis, G. (2014). A review of the effect of traffic and weather characteristics on road safety. Accident Analysis, Prevention, 72, 244-256. https://doi.org/10.1016/j.aap.2014.06.017.
[63] Vilaça, M., Silva, N., Coelho, M. C. (2017). Statistical analysis of the occurrence and severity of crashes involving vulnerable road users. Transportation research procedia, 27, 1113-1120. https://doi.org/10.1016/j.trpro.2017.12.113.
[64] Wang, H., Yeung, D.-Y. (2016). Towards Bayesian deep learning: A framework and some existing methods. IEEE Transactions on Knowledge and Data Engineering 28, 3395-3408. https://doi.org/10.1109/TKDE.2016.2606428.
[65] Wang, K., Zhang, W., Jin, L., Feng, Z., Zhu, D., Cong, H., Yu, H. (2022). Diagnostic analysis of environmental factors affecting the severity of traffic crashes: From the perspective of pedestrian–vehicle and vehicle–vehicle collisions. Traffic injury prevention, 23, 17-22. https://doi.org/10.1080/15389588.2021.1995602.
[66] WAng, J., Mohammed, A. S., Macioszek, E., Ali, M., Ulrikh, D. V., Fang, Q. (2022). A Novel Combination of PCA and Machine Learning Techniques to Select the Most Important Factors for Predicting Tunnel Construction Performance. Buildings, 12, 919. https://doi.org/10.3390/buildings12070919.
[67] Xia, J. S., Khabaz, M. K., Patra, I., Khalid, I., Alvarez, J. R. N., Rahmanian, A., Eftekhari, S. A., Toghraie, D. (2023)a. Using feed-forward perceptron Artificial Neural Network (ANN) model to determine the rolling force, power and slip of the tandem cold rolling. ISA transactions, 132, 353-363. https://doi.org/10.1016/j.isatra.2022.06.009.
[68] Xia, Y., Chen, H., Zimmermann, R. (2023)b. A Random Effect Bayesian Neural Network (RE-BNN) for travel mode choice analysis across multiple regions. Travel behaviour and society, 30, 118-134. https://doi.org/10.1016/j.tbs.2022.08.011.
[69] Xie, Y., Lord, D., Zhang, Y. (2007). Predicting motor vehicle collisions using Bayesian neural network models: An empirical analysis. Accident Analysis, Prevention, 39, 922-933. https://doi.org/10.1016/j.aap.2006.12.014.
[70] 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. https://doi.org/10.1016/j.amar.2019.100109.
[71] Yan, D., Zhou, Q., Wang, J., Zhang, N. (2017). Bayesian regularisation neural network based on artificial intelligence optimisation. International Journal of Production Research, 55, 2266-2287.

https://doi.org/10.1080/00207543.2016.1237785.

[72] Yang, L., Aghaabbasi, M., Ali, M., Jan, A., Bouallegue, B., Javed, M. F., Salem, N. M. (2022). Comparative analysis of the optimized KNN, SVM, and ensemble DT models using Bayesian optimization for predicting pedestrian fatalities: an advance towards realizing the sustainable safety of pedestrians. Sustainability, 14, 10467. https://doi.org/10.3390/su141710467.

[73] Yasmin, S., Eluru, N. (2013). Evaluating alternate discrete outcome frameworks for modeling crash injury severity. Accident Analysis, Prevention, 59, 506-521. https://doi.org/10.1016/j.aap.2013.06.040.

[74] Zajac, S. S., John (2003). Factors influencing injury severity of motor vehicle–crossing pedestrian crashes in rural Connecticut. Accident Analysis, Prevention, 35, 369-379. https://doi.org/10.1016/S0001-4575(02)00013-1.

[75] Zeng, X., Yang, Z., Zhang, L., Tang, X., Zeng, Z., Liu, Z. (2023). Safety verification of nonlinear systems with Bayesian neural network controllers. 15278-15286. https://doi.org/10.1609/aaai.v37i12.26782.

[76] Zhao, B., Zuniga-Garcia, N., Xing, L., Kockelman, K. M. (2023). Predicting pedestrian crash occurrence and injury severity in texas using tree-based machine learning models. Transportation Planning and Technology, 1-22. https://doi.org/10.1080/03081060.2023.2216202.

[77] Zhou, Z.-P., Liu, Y.-S., Wang, W., Zhang, Y. (2013). Multinomial logit model of pedestrian crossing behaviors at signalized intersections. Discrete Dynamics in Nature and Society, 2013. https://doi.org/10.1155/2013/172726.

[78] Zhu, S. (2022). Analyse vehicle–pedestrian crash severity at intersection with data mining techniques. International journal of crashworthiness, 27, 1374-1382. https://doi.org/10.1080/13588265.2021.1929002.