



Article

Prediction of Fatalities at Northern Indian Railways' Road–Rail Level Crossings Using Machine Learning Algorithms

Anil Kumar Chhotu * and Sanjeev Kumar Suman

Civil Engineering Department, National Institute of Technology Patna, Patna 800005, Bihar, India; sksuman@nitp.ac.in

* Correspondence: anilc.phd19.ce@nitp.ac.in

Abstract: Highway railway level crossings, also widely recognized as HRLCs, present a significant threat to the safety of everyone who uses a roadway, including pedestrians who are attempting to cross an HRLC. More studies with new, proposed solutions are needed due to the global rise in HRLC accidents. Research is required to comprehend driver behaviours, user perceptions, and potential conflicts at level crossings, as well as for the accomplishment of preventative measures. The purpose of this study is to conduct an in-depth investigation of the HRLCs involved in accidents that are located in the northern zone of the Indian railway system. The accident information maintained by the distinct divisional and zonal offices in the northern railways of India is used for this study. The accident data revealed that at least 225 crossings experienced at least one incident between 2006 and 2021. In this study, the logistic regression and multilayer perception (MLP) methods are used to develop an accident prediction model, with the assistance of various factors from the incidents at HRLCs. Both the models were compared with each other, and it was discovered that MLP supplied the best results for accident predictions compared to the logistic regression method. According to the sensitivity analysis, the relative importance of train speed is the most important, and weekday traffic is the least important.

Keywords: highway railway level crossing; manned and unmanned level crossing; road user; multilayer perception; logistic regression



Citation: Chhotu, A.K.; Suman, S.K. Prediction of Fatalities at Northern Indian Railways' Road–Rail Level Crossings Using Machine Learning Algorithms. *Infrastructures* **2023**, *8*, 101. <https://doi.org/10.3390/infrastructures8060101>

Academic Editors: Ján Dižo, Rafał Melnik, Stasys Steišūnas and Miroslav Blatnický

Received: 2 April 2023

Revised: 22 May 2023

Accepted: 29 May 2023

Published: 1 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

India has the world's third largest rail network, trailing behind only the United States and China [1]. The magnitude of the Indian rail network is approximately a route of 68,442 km, of which 64,891 km are broad gauge. The rail network serves over 13,500 daily passenger trains (including 5125 suburban EMU trains) and over 9100 daily freight trains [2]. The Indian railway network has recorded 685 crossing accidents between 2006 and 2021 at crosswalks, of which 611 occurred at driverless crossings and 74 at manned crossings, causing 2639 deaths and 4991 non-fatal injuries between 2006 and 2020 [3]. Road traffic accidents account for 43% of all accidents in India [4]. Accidents at driverless crossings during the period 2006–2021 show a downward trend due to some new safety policies by the Indian government. The purpose of this study is to conduct an in-depth investigation of the HRLCs involved in accidents that are located in the northern zone of the Indian Railways system. The records retained by the divisional and zonal offices of the Northern Zone were used to collect the pertinent information, for a total of 225 rail road intersections where at least one accident occurred between 2006 and 2021. This research, unlike some others, investigates a wide variety of factors. Both the structural and functional aspects of crossings are evaluated. Other than the vehicle and train details, numerous different pieces of relevant data, including information such as time, place, driver behaviour, the geometry of crossings, and intersection type, are also included in this study. Results from this study will shed light on the primary factors that contribute to HRLC accidents in the

Northern Railway Vicinity of the Republic of India. Road user safety is extremely crucial at highway and rail intersections. This is because level crossings are areas where two modes of transportation (i.e., rail and motor vehicles) directly interact. This kind of interaction poses significant danger to people and property, and it can result in catastrophic outcomes.

2. Literature Review

Several study efforts have been carried out to date that aim to recognise and examine the numerous reasons for injuries on highways [5–8]. Many researchers have studied the major reasons for accidents at different intersections between roads [9,10]. All such studies have shown that accidents happen at intersections due to the avoidance of traffic. Accidents at HRLCs have been mostly governed by human factors [11]. Some of the common behaviours of humans that result in accidents are overseeding, drunk driving, distractions to the driver, and red light jumping. Avoiding safety gear such as seat belts and helmets, demonstrating a non-adherence to lane driving, and overtaking in the wrong manner are all common reasons for incidence. A human factor analysis regarding rail safety incidents in the United Kingdom was carried out by Madigan et al. [12]. They discovered that operational failures were linked to distractions at work and environmental factors that led to accidents. Das et al. [13] disclosed that fatal accidents occur more often during the day than at night. However, the vehicle's kind and speed are major factors that affect the prevalence of fatal accidents at highway rail grade crossings. Salmon et al. [14] conducted research on the human factors that lead to incidental violations of safety regulations at highway rail grade crossings. Through an investigation into the incident that occurred at the Highways Rail Grade Crossing Khattak and Aleurites, Fordii [15] evaluated pedestrian accidents at highway rail grade crossings and classified them into three levels of severity: "no injury", "injury", and "fatality". The results demonstrated that pedestrians are also vulnerable to fatalities at highway-rail-grade crossings due to the higher train speeds. Liu et al. [16] evaluated pre-crash driver behaviour at highway-rail-grade crossings with differing types of warning devices, and the outcomes showed that drivers were likely to stop at HRCGs with gates. Flashing lights and perceptible warning devices at gates were found to be effective means of safety at crossings. Larue et al. [17] conducted an analysis of the threats and misjudgements that could be made by motorists who use highways and rail grade crossings that are fully protected. Numerous violations by motorists and pedestrians were observed at the proposed highway-rail-grade crossing. It has been suggested that planning issues can lead to an increase in violations at fully operational highway-rail-grade crossings. Keramati et al. [18] considered the effect of different geometric parameters on accidents at HRLCs. Consideration was given to the acute crossing angle, width (proportional to the number of tracks), distance between the highway rail grade crossing and the nearest intersection, and the number of lanes on the highway. The results of the conducted research demonstrated that all the considered geometric characteristics can significantly affect accident severity and occurrence. From the above study, we found that many factors, such as human, environmental, seasonal, and geometrical ones, and these factors influenced the accident rate at HRLCs. Many studies have been carried out to establish a relationship between different factors in accidents by using different mathematical and statistical tools such as logistic regression, poison distribution, and binomial distribution [19]. Apart from this, some of the latest soft computing technologies are used in the prediction of road accidents. ANN is used in the transportation sector to enhance mobility and safety [20,21]. Xie et al. [22] compared the Bayesian artificial neural network (ANN), ANN (backpropagation), and negative binomial regression modelling techniques. They also revealed that the ANN and Bayesian ANN models significantly outperformed negative binomial regression in predicting traffic accidents. Najjar et al. [23] implemented a back-propagation ANN for establishing the speed limits on two-lane highways in Kansas. They utilized four roadway-related input parameters: shoulder width, shoulder type, ADT, and the percentage of no-passing zones. In addition to predicting 85th percentile speeds, the ANN model was created to predict the potential effects of changes to specific

roadway and traffic-related parameters. The developed ANN was able to predict the 85th percentile speed with an average degree of accuracy of approximately 96%. In order to analyse and predict traffic accidents in Sudan, Ali and Bakheit [24] used an ANN (artificial neural network) model. ANN models include principal component regression models. The results show that ANN models fit the data more closely (as measured by the coefficient of determination), but the predictions are otherwise very similar. Delen et al. [25] used police reports of 30,358 car accidents between 1995 and 2000 to create eight binary multilayer perceptron (MLP) neural network models, with different levels of injury (from no injury to death) as the dependent variable. Their model helps one to find the most important factors that explain each dependent variable. By using an ANN, Jadaan et al. [26] developed a model for predicting incidences by looking at the relationship between injuries and the factors that affect them. The model produced results that were good for Jordanian traffic. Alkheder et al. [27] trained an ANN model to predict the severity of injury (mild, moderate, severe, and fatal) of avenue visitor accidents with the data from 5973 incidents that occurred in Abu Dhabi between 2008 and 2013. Overall, their model predicted an average success rate of 74.6 percent when using the testing dataset. Sameen and Pradhan [28] made a recurrent neural network (RNN) to be ready for different kinds of injuries. The RNN version was compared to the MLP and Bayesian logistic regression models (BLR). The RNN version was found to be more correct than the ANN and BLR versions. Borja et al. [29] offered a method for founding an accident threat prediction model. They developed models with artificial neural networks (ANNs) and decided on the ultimate structure of the ANN version, which enabled the use of information for incidence counts on the Swiss national roads (2009–2012). It becomes clear that ANNs may be used as a workable approach to predict the frequency of street accidents. In addition to the above, the emergence of various datasets has led to the use of different prediction methods for various engineering problems [30–34]. Pattern recognition equipment and the correct evaluation of its usage of optimized prediction obligations are contemporary subject matters in current years [35–39]. From the above literature study, it was concluded that HRLC accidents occur due to various factors. To establish a relationship between these factors, we need a mathematical tool such as logistic regression and ANN, which are widely used in accident prediction; however, very little research has been carried out on HRLC accident prediction. From the literature review, it was also found that most of the study was conducted in developed economies where advanced and intelligent infrastructure is available. Apart from this, the education level of the people was also high, due to which people can understand the importance of safety and its effects more wisely. Hence, this study concentrated on those developing economies where a lack of advanced infrastructure and low education levels can differ from the results of the previous study.

3. Methodology

3.1. Selection of Study Area and Data Collection

This study collected data from the Northern Railways, North Central Railways, North Western Railways, North Eastern Railways, and Northeast Frontier Railways in India. The total length of railways in the abovementioned railway zone is 23,319 km [40], and this encompasses different parts of different Indian states. Data were collected from the database of the zonal head office and divisional head office via the right to information (RTI) act and from a direct visit to the office. The accident data were collected from 2006 to 2021. These data contain the following: place of accident, time, date, type of train involved, type of vehicle involved, number of fatalities and injuries, type of injuries, and manned or unmanned level crossing. The sample demonstration of the dataset is shown in Table 1.

Table 1. Sample demonstration of the dataset.

S. No.	Date of Accident	Brief Description	Casualties			Reason
			Killed	Major Injuries	Minor Injuries	
1	21 January 2014—01:35	Train No.12,485 Up Nanded-Sri Ganganagar Express left Pakki at 01:23 hr towards Abohar. While the train was approaching Manned Level Crossing Gate No A/47-A (Engineering, Interlocked Gate) between Pakki and Abohar stations, one Car (No. PB-10DW-7202, Toyota Etios Liva), after hitting the closed boom of MLC Gate No. A-47/A, dashed against the train engine, thus causing the death of 02 car occupants. The car driver was unhurt.	2	0	0	Negligent driving by a road vehicle driver who did not stop at the closed gate.
2	9 December 2012—18:48	Maruti car no-PB-08W-1789 was stuck with train no-54,621 at manned level crossing gate no-A-82 between the Dasua-Khudda Kurala part of the Pathankot-Jalandhar section.	2	0	1	L-xing Gate A-82, before granting a line clear to train No-54,621 to Station Master/Khuda Kurala (due to which the gate remained in an open condition), resulted in an accident.

3.2. Primary Investigation of the Accident Data of Northern Railways

This study primarily concentrates on looking at the characteristics of RRLCs that experienced accidents in the northern zone of Indian Railways between 2006 and 2021. A total of 225 crossings were found in the northern zone out of 355 unmanned level crossings, where at least one accident occurred between 2006 and 2021. The data illustrate that the number of RRLC accidents increased from 2006 to 2014, then decreased over the next seven years. From the data, it is observed that fatalities were highest in 2011, whereas 2019 and 2020 had no accidents at level crossings. The drastic decrease in accidents is due to some of the major safety enhancement policies and planning in road safety that has been conducted by the Government of India. By 2025, the Indian government intends to remove 2500 unmanned level crossings from national highways [41]. The majority of level crossings are regularly maintained. The primary objective of the Indian government is to improve the existing infrastructure of railways through the routine monitoring of level crossings, road signs and signals, and surface types. Another cause for the reduction in accidents in 2020–21 was the lockdown that occurred due to the spread of COVID-19. Indian Railways is planning to remove all unmanned levels from major national highways by the year 2022, which is another reason for accident reduction [40]. Railroad level crossing (RRLC) casualties in the northern zone depend on the type of crossing, the presence of light, the surface of the intersection area, the type of warning system deployed, traffic characteristics,

driver characteristics, and environmental factors. There is a total of 87 RRLCs at the most-threatened crossings, which is where the lighting is inadequate. Between 2006 and 2021, non-gated RRLCs accounted for 86.7% of accidents. Most of the crossings have crossbucks or stop signs that are not properly maintained, some road signs that are broken, and some that are found to be faded in colour. According to the data, 20% of RRLCs have inadequate protection, which is one of the causes of accidents. According to Figure 1, the majority of accidents (88%) occur during the day because trains and road traffic interact more during the day than at night. Countries such as India that have a daytime work culture are also a prominent cause for daytime accidents. Most of the accidents that occur at unmanned level crossings compared to manned level crossings are shown in Figure 2.

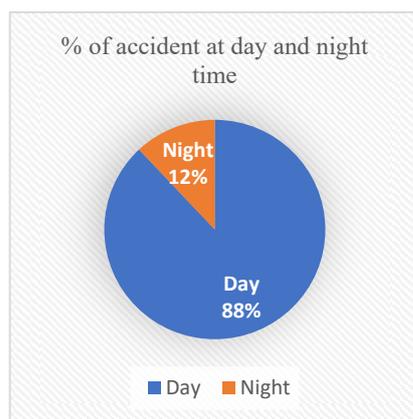


Figure 1. Accident at day and night times.

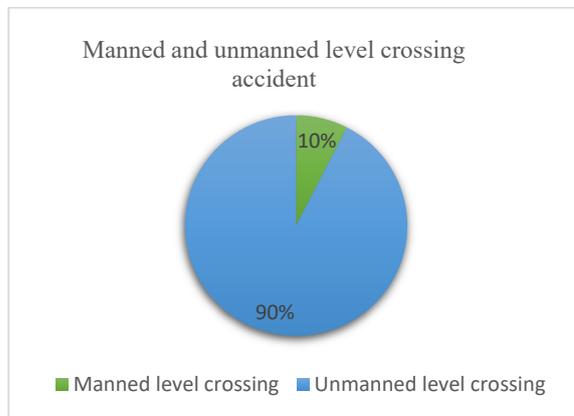


Figure 2. Accident at manned and unmanned level crossings.

These events occurred because unmanned level crossings are not protected by gates. In the study area, several unmanned level crossings were found without proper signs, broken stop signs, and several had road signs that had faded in colour.

Passenger trains are more involved in accidents compared to goods trains, as shown in Figure 3. This is because goods trains move at a slower speed compared to passenger trains. In the northern zone of railways, the number of passenger trains is higher compared to goods trains, so there is less interaction between vehicle and train (which is a major reason for passenger train accidents). Accidents at level crossings are also influenced by the geometry of the crossings. More accidents occur at skewed-geometry level crossings than at linear ones. This is because the motorist has less available sight distance.

In the study area, the trains run from major cities where traffic volume is very high; as such, more train-traffic interaction takes place. Due to this, more accidents occur in urban areas compared to rural areas, as shown in Figure 4. The dry season has fewer accidents than the wet season. The dry season is summer, while the wet season is autumn and winter.

In the wet season, visibility is disturbed by heavy rain and fog. For the study area, a similar result shows up in Figure 5. Peak hours experience more accidents compared to non-peak hours due to the increased interaction of vehicles and trains. This is also valid for the study area, as shown in Figure 6.

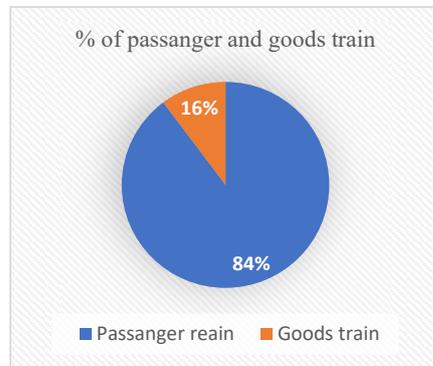


Figure 3. Type of trains involved in accidents.

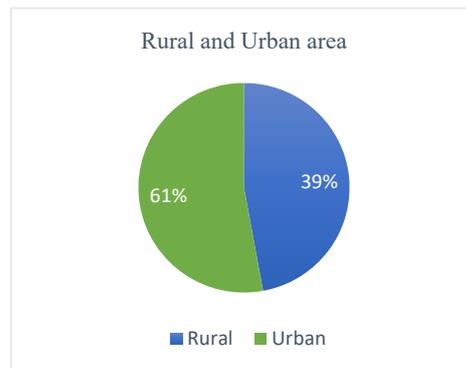


Figure 4. Accidents in rural and urban areas.

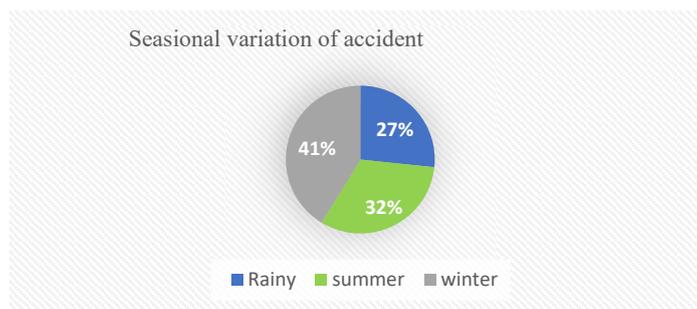


Figure 5. Seasonal variations in accident.

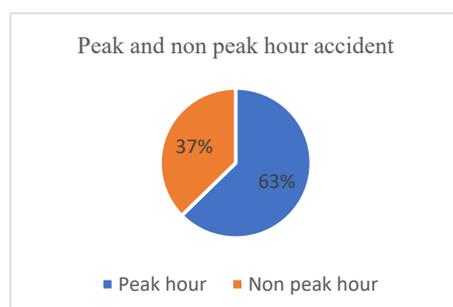


Figure 6. Peak and non-peak hour accidents.

Descriptive Statistics of the Variable

In descriptive statistics, the maximum, minimum, mean, standard deviation, and variance of all variables are calculated and tabulated in Table 2. The speed of the train varies from 22 km/h to 120 km/h. However, the average train speed is 64.9 km/h. The variation in speed is also shown in Figure 7.

Table 2. Descriptive statistics of the variable.

Variable	Descriptive Statistics					
	N	Min.	Max.	Mean	Std. Deviation	Variance
Rural or urban area (A_{UR})	225	0	1	0.64	0.480	0.230
Fatal and non-fatal accidents (A_{FN})	225	0	1	0.631	0.483	0.234
No. of railway track (T_N)	225	0	1	0.61	0.488	0.238
Day and night (T_{DN})	225	0	1	0.573	0.495	0.246
Weather (W_{DW})	225	0	1	0.587	0.497	0.244
Manned and unmanned level crossings (LC_{MU})	225	0	1	0.827	0.380	0.144
Surface type (S_{BC})	225	0	1	0.462	0.499	0.250
Average speed (V)	225	22.0	120.0	64.9	23.9	571.0
Type of train (T_{PG})	225	0	1	0.740	0.438	0.192
Vehicle type (V_{CN})	225	0	1	0.710	0.464	0.235
Road geometry (G_{CS})	225	0	1	0.524	0.505	0.251
Warning device (W_{IN})	225	0	1	0.58	0.495	0.551
Weekend and weekdays (W_{WWD})	225	0	1	0.267	0.443	0.196
Peak and non-peak hours (H_{PN})	225	0	1	0.360	0.481	0.231
Gauge of track (G_{BM})	225	0	1	0.733	0.434	0.0197
Negligence of driver or gateman (N_{GD})	225	0	1	0.667	0.472	0.223

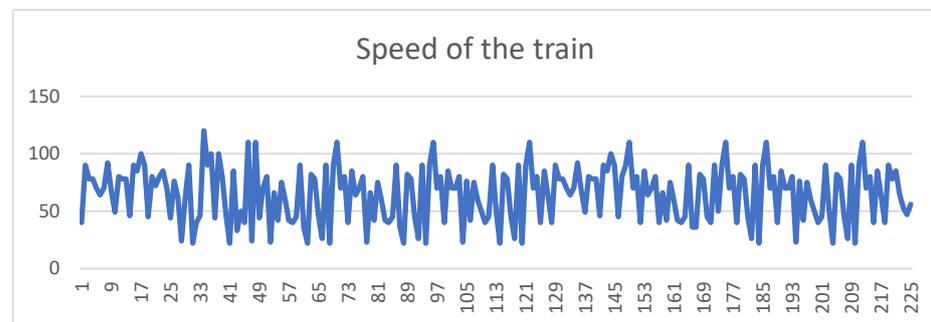


Figure 7. Variation in the speed of trains.

3.3. Model Development and Analysis

3.3.1. Models

There are two methods that were used for the analysis of data in this study. Analysis was completed by using both methods, and the results were compared.

- I. Logistic regression;
- II. Artificial neural network.

Logistic Regression

The regression method, also known as logistic regression, was used to fit the accident data. In order to predict future events, probabilistic systems were modelled using logistic regression techniques. The distributions of the explanatory variables or predictors were not necessary in these direct probability models [41]. If p is the probability that a binary response variable $Y = 1$ when input variable $X = x$, then the logistic response function is modelled as

$$P = P(Y = 1|X = x) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n}} \tag{1}$$

This function represents an s-shaped curve and is non-linear. Here, β is the coefficient of the predictor or input of the variable x that is used in a regression equation.

Artificial Neural Networks

A neural network machine-learning model has been extensively used in predictive applications. Warren McCulloch, a neurophysiologist, and Walter Pits, a logician, based the first artificial neuron on a biological neuron in 1943 [42]. In artificial neural networks, feedforward networks and feedback networks are the two main architecture types. Feedforward, or multi-layer, networks have been used quite often when constructing neural models. In such models, several layers as a hidden layer and one output layer may be included. The general mathematical expression of an ANN model is represented in Equation (2).

$$A_N = \varphi \left[b_o + \sum_{k=1}^{k=m} \left\{ w_k * \varphi \left(b_k + \sum_{i=1}^{i=n} w_{ik} x_i \right) \right\} \right] \tag{2}$$

where A_N = normalized output of the model; φ = activation function, b_o = bias at the output layer neuron; w_k = weight between the output layer neuron and k th neuron of the hidden layer; b_k = bias associated to k th neuron of the hidden layer; w_{ik} = weight between i th neuron of the input layer and k th neuron of the hidden layer; x_i = normalized i th variable (neuron) of the input layer; n = number of input variables; and m number is the neurons in the hidden layer.

3.3.2. Preparation of Model Data

In this paper, in order to establish a predictive model for railroad level crossing accidents, fatal and non-fatal accidents were selected as the dependent variables. A fatal accident is coded as $y = 1$ and a non-fatal accident as $y = 0$. Another variable is shown in Table 1 with a coded value. All dependent variables were coded as shown in Table 3.

Table 3. Details of the independent variables for the proposed model.

Variable	Abbreviation of Variables	Measure of Variable	Coded Value
Rural or urban area	A _{UR}	Nominal	0 = Rural area, 1 = urban area
Fatal/non-fatal accidents	A _{FN}	Nominal	0 = Non-fatal, 1 = Fatal
No. of railway track	T _N	Nominal	0 = One track, 1 = For two-track
Day and night	T _{DN}	Nominal	0 = Day time, 1 = Night time

Table 3. *Cont.*

Variable	Abbreviation of Variables	Measure of Variable	Coded Value
Weather	W_{DW}	Nominal	0 = Dry weather, 1 = Wet weather
Manned and unmanned level crossings	LC_{MU}	Nominal	0 = Manned level crossing, 1 = Unmanned level crossing
Road surface type	S_{CE}	Nominal	0 = Concrete, 1 = Earthen
Average speed	V	Nominal	0 = less than 50, 1 = greater than 50
Type of train	T_{PG}	Nominal	0 = Passenger train, 1 = Goods train
Vehicle type	V_{LH}	Nominal	0 = Light vehicle, 1 = Heavy vehicle
Road geometry	G_{CS}	Nominal	0 = Curve, 1 = Straight
Warning device	W_{IN}	Nominal	0 = Not installed properly, 1 = Installed properly
Weekend and weekdays	W_{WWD}	Nominal	0 = Weekend, 1 = Weekdays
Peak and non-peak hours	H_{PN}	Nominal	0 = Non peak hour, 1 = Peak hour
Gauge of track	G_{BM}	Nominal	0 = Meter gauge, 1 = Broad gauge
Negligence of driver or gateman	N_{GD}	Nominal	0 = Gateman, 1 = Driver

4. Result

4.1. Result of Logistic Regression Model

The binary logistic regression model that was built within an IBM SPSS Statistics 22 environment was used for the analysis. The findings of the statistical analysis are summarized in Table 4, including the following information for each predictor: (1) estimate (2) standard error; (3) Wald; (4) degree of freedom; (5) *p*-value; and (6) Exp (B).

According to the findings of the statistical analysis, the majority of the considered predictor variables were statistically significant with *p*-values of less than 0.05. (See Table 4). Furthermore, for the entire model, the *p*-value was less than 0.001, which shows that the model was statistically significant. Moreover, some of the predictor variables considered had relatively high *p*-values. This is because certain aspects were shared by all the RRLCs that had at least one accident during the 15-year study period. The road surface at crossings was not significant because the crossing is a very short distance; thus, it does not have much impact. The majority of the manned RRLCs in the northern zone of the Indian railways have active warning devices installed. The *p*-value for gauge of track was 0.210, which is not statistically significant because the majority of RRLCs in the northern zone have similar gauges of trains.

Table 4. Results of the logistic regression using fatality as the dependent variable.

Variable	Estimates	S.E.	Wald	df	p-Value
Rural or urban area	8.941	2.525	12.541	1	0.000
No. of railway track	6.794	2.526	7.234	1	0.007
Day and night	3.823	1.492	6.567	1	0.010
Weather	3.067	1.720	3.179	1	0.045
Manned and unmanned level crossings	−1.233	1.594	0.599	1	0.042
Road surface type	−1.185	1.309	0.820	1	0.365
Average speed	0.237	0.068	12.156	1	0.000
Type of train	−1.725	1.928	0.800	1	0.371
Vehicle type	−0.716	0.587	1.487	1	0.223
Road geometry	−0.640	1.351	0.225	1	0.047
Warning device	2.320	1.235	3.531	1	0.048
Weekend and weekdays	4.119	2.213	3.464	1	0.063
Peak and non-peak hours	0.744	1.354	0.301	1	0.583
Gauge of track	0.271	1.400	0.117	1	0.847
Negligence of driver or gateman	−0.444	1.295	0.037	1	0.032
Interceptions	−27.037	9.381	8.306	1	0.004

The area under the curve (AUC) for the MLP model is 0.94, which is more than 0.90; hence, the model can distinguish between fatal and non-fatal accidents very well. The accuracy of the model is 0.97, which is close to 1.0; this implies that the model can accurately predict fatalities 97 times out of 100 with the given condition of level crossings. This is shown in Table 5.

Table 5. Confusion matrix for the logistic regression.

Model	Confusion Matrices		Accuracy	Sensitivity	Specificity	AUC	
	Non-fatal	Fatal					
Logistic regression	Non-fatal	80	3	0.96	0.98	0.09	0.94
	Fatal	2	140				

Logistic Regression Model Validation

Four pseudo-R-square statistical tests were used to assess the fitness of the proposed model, giving satisfactory results for all tests. First, the -2 log-likelihood (or -2LL) test was performed (this is also referred to as the model deviance). The lowest value of the -2LL was zero, which signifies a perfect predictive performance (increasing values relative to zero indicate a worse model fit [43]). This indicator is typically not very insightful regarding the characteristics of a poor fitted model. For the proposed model, the value of -2LL was 0.089, which is near zero and indicates that the model is fit for prediction. The second test used was Cox and Snell’s R square. Cox and Snell’s R-square is based on the log likelihood of the model compared to the log likelihood of a baseline model. However, with categorical outcomes, it has a theoretical maximum value of less than 1, even for a “perfect” model. The most significant value of the Cox and Snell R square is 1, which indicates a perfect fit, and decreasing values relative to 1 signify a worse model fit [43]. For the proposed model, this value was 0.943, which is near one and signifies a good fit. The third statistical test performed was the Nagelkerke R-square. The largest value of the Nagelkerke R square is 1, which indicates a perfect fit, and decreasing values relative to 1 signify a worse model fit [44]. For the proposed model, this value was 0.973, which is near one and signifies a

good fit. The fourth test used was McFadden’s pseudo-R-square, which lies between 0.2 and 0.4 for a good-fit model. For the proposed model, this value was 0.31, which shows that the model is a perfect fit. The Hosmer and Lemeshow [45] test provide an additional global fit test, comparing the estimated model to one with a perfect fit. If this assessment is not significant, it reveals that the model is a well-specified fit model. If it is significant, then we have evidence that the model is misspecified or does not fit the model. Here, the Hosmer and Lemeshow tests were not statistically significant [$\chi^2(8) = 0.286, p = 1.000$], suggesting that the model fits adequately. From the above details, it was found that the proposed model satisfied all statistical fitness tests.

4.2. Results of the ANN Model

The ANN model used in this study was developed using 15 independent variables, as per Table 2, by taking fatal and non-fatal accidents as the dependent variables. Optimization of the model was done with the gradient descent algorithm. The activation function for the input layer and output layers were hyperbolic tangent and sigmoid, and these give a maximum accuracy of the model, as shown in Table 4. The accuracy of training and testing was 100% for fatal accident prediction. Non-fatal accident accuracy for training and testing were 96.9 and 94.7, respectively. There are seven hidden layers used in this model. The confusion matrix for the ANN model is shown in Table 6.

Table 6. Results of the ANN model.

Model	Activation Function		Confusion Matrices				Accuracy	
			Training		Testing		Training	Testing
MLP Model	Input t Layer	Output Layer	Fatal	Non Fatal	Fatal	Non Fatal		
			Fatal	0	44	0	100	100
			Non-fatal	2	62	1	18	96.9

4.2.1. Area under Curve (AUC) from ROC Curve

The area under the curve (AUC) for the MLP model is 0.986, which is more than 0.90; hence, the model can distinguish between fatal and non-fatal accidents very well [46]. This is calculated using the ROC curve, which is drawn using specificity on the X axis and sensitivity on the Y axis.

4.2.2. Sensitivity Analysis for the ANN Model

A sensitivity analysis was carried out in order to determine which of the many possible factors had the greatest impact. The connection weights in a neural network model were deduced using the formulas proposed by Garson [47]. In addition to that, Shahin et al. [48] applied this theory to the field of civil engineering. We were able to determine the relative importance (RI) of each independent variable with the assistance of this analysis. According to Table 7, the variables were ranked based on the decreasing order of their corresponding relative importance values. As the ranks were assigned to each of the variables in Table 6, it was observed that the variable ‘speed of the train’ had the highest RI of 32.1%. The ‘level crossing type’ variable was observed to be the second most important variable. Similarly, “weekend and weekend day traffic” was discovered to have the lowest impact on model output.

Table 7. Sensitivity analysis of all variables.

Variable	T _R	A _{UR}	S _{SW}	D _{DN}	L _{CMU}	W _{WWD}	V	W _{YN}	G _{CS}	H _{PN}	N _{DR}	G _{BM}	S _{CE}
RI	6.1	7.2	6.3	8.2	9.7	4.0	32.1	5.4	4.6	9.4	4.3	1.2	1.5
Rank	7	5	6	3	2	11	1	8	9	4	10	13	12

5. Discussion

In urban India, especially in large cities, concerns have been raised about the security of vehicle users at level crossings and other such intersections. While statistical analysis and modelling have been widely used to assess pedestrian safety at traffic junctions [49], there is a need for a more holistic approach. This research has presented a new approach and is based on the use of logistic regression and artificial neural network techniques for determining the connection between level crossing characteristics and fatal and nonfatal accidents. In this investigation, efforts were made to construct ANN-based models for determining the frequency of fatal pedestrian collisions. Nevertheless, several researchers have elucidated the importance of statistical models for predicting accident frequencies [50,51]. Chakraborty and Mitra [52] created a negative binomial model to predict the catastrophic pedestrian accident frequency in Kolkata; they demonstrated that the statistical model’s prediction performance was nearly 50%. However, the accuracy of an ANN model is very high when compared to statistical models [53–56]. Alkheder et al. [27] used an ANN model to predict the degree of injury (minor, moderate, severe, and death) of avenue visitor accidents. Their version had an average overall prediction performance of 74.60%. In this investigation, the accuracy level for logistic regression was 96%, but the accuracy level for the ANN model that uses hyperbolic tangent as its activation function for the input layer and SoftMax for the output layer was 98%. Both models have better accuracy than the statistical models used by various researchers. In addition to analysing data on the frequency of road–railway events, it is also important to analyse the data on the triggers that contribute to fatalities and the extent to which these factors impact. Most of the previous studies where ML was used as an effective factor in accidents were smaller than the present study. Some of the important factors, such as average daily traffic, age of the driver, sight distance, and frequency of trains, may be included in future studies. These data are not available, but they are very important for accident prediction.

6. Conclusions

In this study, the accident data from the past 15 years on rail road level crossings in the northern zone of Indian railways were analysed. From the data, it was found that level crossing accidents have decreased due to some initiatives by Indian railways. Unmanned level crossings encounter more accidents compared to manned level crossings. Most of the unmanned level crossings have road markings and signs that are either faded or not properly installed. The speed of the trains, day and night driving, weather, rural and urban areas, the number of railway tracks, and the surface type of the pavement at highway–rail-grade crossings were found to be the factors significantly affecting the severity of driver injuries at both manned and unmanned level crossings. Some of the factors, such as the availability of signboards, road markings, and average annual daily traffic, were not significant for the prediction of accidents, as per the proposed model. Multilayer perception ANN has an accuracy of 98%, while logistic regression has an accuracy of 97%. As per the sensitivity analysis, the speed of train had the greatest relative impact (32.1) on accidents; moreover, the gauge of track had the least relative importance (1.2). This paper includes only 15 independent variables, but some of the variables may be included in future work, such as driver and pedestrian behaviours, sight distance, delays at the level crossing, automatic and manual gate operations, and the width of the road near crossings. In Indian conditions, fatal and non-fatal accidents at RRLCs can be reduced by increasing driver and pedestrian awareness and by improving safety standards. The government must impose severe penalties on drivers who violate traffic laws at intersections such

as RRLCs. Intelligent signal management and monitoring systems must be adopted for the effective reduction in accidents at level crossings. Recommendations are made for addressing traffic engineering, road, and construction concerns to enhance the security of road–railway infrastructure. In addition, creating and enforcing more stringent laws, particularly regarding identified causes of fatal accidents, and increasing penalties is recommended. State and federal governments should set aside money for the development of national and local databases that compile data on road–railway collisions, including the frequency of rail links, the average age of drivers, the levels of education and income they hold, the types of property damage that they sustain in accidents, and more.

Author Contributions: A.K.C. and S.K.S.: conceptualization, methodology, formal analysis, investigation, and resources; A.K.C.: software, validation writing, visualization, and original draft; S.K.S.: supervision and review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Acknowledgments: We would like to express our most sincere heartfelt gratitude and respect to our respected supervisor, Dr. Sanjeev Kumar Suman, Associate Professor, National Institute of Technology, Patna, for his excellent guidance, valuable suggestions, and endless support throughout the work. We would like to express gratitude and sincere thanks to Prof. S. S. Mishra, Head of the Department of Civil Engineering. We are greatly indebted to them for their constructive suggestions and criticism during the progress of the work. Finally, we express our deepest gratitude to our family and friends for their continuous encouragement, understanding, and support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Welcome to Indian Railway Passenger Reservation Enquiry. Available online: <https://www.indianrail.gov.in/enquiry/StaticPages/StaticEnquiry.jsp?StaticPage=index.html> (accessed on 25 September 2022).
2. Annual Statistical Statement 2018–19—Indian Railway. Available online: https://indianrailways.gov.in/railwayboard/view_section.jsp?lang=0&id=0,1,304,366,554 (accessed on 10 October 2022).
3. Accidental Deaths & Suicides in India (ADSI) | National Crime Records Bureau. Available online: <https://ncrb.gov.in/en/accidental-deaths-suicides-india-ads-i> (accessed on 22 October 2022).
4. Road Accidents in India. Available online: <https://www.statista.com/topics/5982/road-accidents-in-india/> (accessed on 1 November 2022).
5. Clifton, K.J.; Kreamer-Fulfs, K. An Examination of the Environmental Attributes Associated with Pedestrian–Vehicular Crashes near Public Schools. *Accid. Anal. Prev.* **2007**, *39*, 708–715. [[CrossRef](#)] [[PubMed](#)]
6. Dai, D. Identifying Clusters and Risk Factors of Injuries in Pedestrian–Vehicle Crashes in a GIS Environment. *J. Transp. Geogr.* **2012**, *24*, 206–214. [[CrossRef](#)]
7. Mrema, I.J.; Dida, M.A. A Survey of Road Accident Reporting and Driver’s Behavior Awareness Systems: The Case of Tanzania. *Eng. Technol. Appl. Sci. Res.* **2020**, *10*, 6009–6015. [[CrossRef](#)]
8. Mohan, D.; Tsimhoni, O.; Sivak, M.; Flannagan, M.J. *Road Safety in India: Challenges and Opportunities*; University of Michigan, Ann Arbor, Transportation Research Institute: Ann Arbor, MI, USA, 2009.
9. Mohan, D.; Tiwari, G.; Mukherjee, S. Urban traffic safety assessment: A case study of six Indian cities. *IATSS Res.* **2016**, *39*, 95–101. [[CrossRef](#)]
10. Muley, D.; Kharbeche, M.; Alhajyaseen, W.; Al-Salem, M. Pedestrians’ crossing behavior at marked crosswalks on channelized right-turn lanes at intersections. *Procedia Comp. Sci.* **2017**, *109*, 233–240. [[CrossRef](#)]
11. Railroad Accidents: Common Causes, Statistics and Prevention. Available online: <https://www.sidgilreath.com/learn/railroad-accidents-causes.html> (accessed on 10 January 2023).
12. Madigan, R.; Golightly, D.; Madders, R. Application of Human Factors Analysis and Classification System (HFACS) to UK Rail Safety of the Line Incidents. *Accid. Anal. Prev.* **2016**, *97*, 122–131. [[CrossRef](#)]
13. Das, S.; Kong, X.; Lavrenz, S.M.; Wu, L.; Jalayer, M. Fatal Crashes at Highway Rail Grade Crossings: A U.S. Based Study. *Int. J. Transp. Sci. Technol.* **2022**, *11*, 107–117. [[CrossRef](#)]
14. Salmon, P.M.; Read, G.J.M.; Stanton, N.A.; Lenné, M.G. The Crash at Kerang: Investigating Systemic and Psychological Factors Leading to Unintentional Non-Compliance at Rail Level Crossings. *Accid. Anal. Prev.* **2013**, *50*, 1278–1288. [[CrossRef](#)]
15. Khattak, A.; Tung, L.-W. Severity of Pedestrian Crashes at Highway-Rail Grade Crossings. *J. Transp. Res. Forum* **2015**, *54*, 91–100. [[CrossRef](#)]

16. Liu, J.; Khattak, A.J.; Richards, S.H.; Nambisan, S. What Are the Differences in Driver Injury Outcomes at Highway-Rail Grade Crossings? Untangling the Role of Pre-Crash Behaviours. *Accid. Anal. Prev.* **2015**, *85*, 157–169. [CrossRef]
17. Larue, G.S.; Naweed, A.; Rodwell, D. The Road User, the Pedestrian, and Me: Investigating the Interactions, Errors, and Escalating Risks of Users of Fully Protected Level Crossings. *Saf. Sci.* **2018**, *110*, 80–88. [CrossRef]
18. Keramati, A.; Lu, P.; Tolliver, D.; Wang, X. Geometric Effect Analysis of Highway-Rail Grade Crossing Safety Performance. *Accid. Anal. Prev.* **2020**, *138*, 105470. [CrossRef] [PubMed]
19. Moodie, E.E.M. A Review of: “An Introduction to Generalized Linear Models, Third Edition, by A. J. Dobson and A. G. Barnett”. *J. Biopharm. Stat.* **2009**, *19*, 307. [CrossRef]
20. Xu, C.; Tarko, A.P.; Wang, W.; Liu, P. Predicting Crash Likelihood and Severity on Freeways with Real-Time Loop Detector Data. *Accid. Anal. Prev.* **2013**, *57*, 30–39. [CrossRef] [PubMed]
21. Sohn, S.Y.; Shin, H. Pattern Recognition for Road Traffic Accident Severity in Korea. *Ergonomics* **2001**, *44*, 107–117. [CrossRef] [PubMed]
22. Xie, Y.; Lord, D.; Zhang, Y. Predicting Motor Vehicle Collisions Using Bayesian Neural Network Models: An Empirical Analysis. *Accid. Anal. Prev.* **2007**, *39*, 922–933. [CrossRef]
23. Najjar, Y.M.; Stokes, R.W.; Russell, E.R. Setting Speed Limits on Kansas Two-Lane Highways: Neuronet Approach. *Transp. Res. Rec.* **2000**, *1708*, 20–27. [CrossRef]
24. Ali, G.A.; Bakheit, C.S. Comparative analysis and prediction of traffic accidents in Sudan using artificial neural networks and statistical methods. In Proceedings of the 30th South African Transport Conference, Centurion, South Africa, 11–14 July 2011; Document Transformation Technologies: Centurion, South Africa, 2011; pp. 202–214.
25. Delen, D.; Sharda, R.; Bessonov, M. Identifying Significant Predictors of Injury Severity in Traffic Accidents Using a Series of Artificial Neural Networks. *Accid. Anal. Prev.* **2006**, *38*, 434–444. [CrossRef]
26. Jadaan, K.S.; Al-Fayyad, M.; Gammoh, H.F. Prediction of Road Traffic Accidents in Jordan Using Artificial Neural Network (ANN). *J. Traffic Logist. Eng.* **2014**, *2*, 92–94. [CrossRef]
27. Alkheder, S.; Taamneh, M.; Taamneh, S. Severity Prediction of Traffic Accident Using an Artificial Neural Network. *J. Forecast.* **2016**, *36*, 100–108. [CrossRef]
28. Sameen, M.; Pradhan, B. Severity Prediction of Traffic Accidents with Recurrent Neural Networks. *Appl. Sci.* **2017**, *7*, 476–492. [CrossRef]
29. Garcia de Soto, B.; Bumbacher, A.; Deublein, M.; Adey, B.T. Predicting Road Traffic Accidents Using Artificial Neural Network Models. *Infrastruct. Asset Manag.* **2018**, *5*, 132–144. [CrossRef]
30. Wang, B.; Zhang, L.; Ma, H.; Wang, H.; Wan, S. Parallel LSTM-Based Regional Integrated Energy System Multienergy Source-Load Information Interactive Energy Prediction. *Complexity* **2019**, *2019*, 7414318. [CrossRef]
31. Alshboul, O.; Shehadeh, A.; Almasabha, G.; Mamlook, R.E.A.; Almuflih, A.S. Evaluating the Impact of External Support on Green Building Construction Cost: A Hybrid Mathematical and Machine Learning Prediction Approach. *Buildings* **2022**, *12*, 1256. [CrossRef]
32. Singh, P.; Pasha, J.; Moses, R.; Sobanjo, J.; Ozguven, E.E.; Dulebenets, M.A. Development of Exact and Heuristic Optimization Methods for Safety Improvement Projects at Level Crossings under Conflicting Objectives. *Reliab. Eng. Syst. Saf.* **2022**, *220*, 108296. [CrossRef]
33. Alshboul, O.; Almasabha, G.; Shehadeh, A.; Mamlook, R.E.A.; Almuflih, A.S.; Almakayeel, N. Machine Learning-Based Model for Predicting the Shear Strength of Slender Reinforced Concrete Beams without Stirrups. *Buildings* **2022**, *12*, 1166. [CrossRef]
34. Zheng, S.; Lyu, Z.; Foong, L.K. Early Prediction of Cooling Load in Energy-Efficient Buildings through Novel Optimizer of Shuffled Complex Evolution. *Eng. Comput.* **2020**, *38*, 105–119. [CrossRef]
35. Zhu, W.; Ma, C.; Zhao, X.; Wang, M.; Heidari, A.A.; Chen, H.; Li, C. Evaluation of Sino Foreign Cooperative Education Project Using Orthogonal Sine Cosine Optimized Kernel Extreme Learning Machine. *IEEE Access* **2020**, *8*, 61107–61123. [CrossRef]
36. Liu, G.; Jia, W.; Wang, M.; Heidari, A.A.; Chen, H.; Luo, Y.; Li, C. Predicting Cervical Hyperextension Injury: A Covariance Guided Sine Cosine Support Vector Machine. *IEEE Access* **2020**, *8*, 46895–46908. [CrossRef]
37. Kozłowski, E.; Borucka, A.; Świdorski, A.; Skoczyński, P. Classification Trees in the Assessment of the Road–Railway Accidents Mortality. *Energies* **2021**, *14*, 3462. [CrossRef]
38. Tang, H.; Xu, Y.; Lin, A.; Heidari, A.A.; Wang, M.; Chen, H.; Luo, Y.; Li, C. Predicting Green Consumption Behaviors of Students Using Efficient Firefly Grey Wolf-Assisted K-Nearest Neighbor Classifiers. *IEEE Access* **2020**, *8*, 35546–35562. [CrossRef]
39. Shehadeh, A.; Alshboul, O.; Al Mamlook, R.E.; Hamedat, O. Machine Learning Models for Predicting the Residual Value of Heavy Construction Equipment: An Evaluation of Modified Decision Tree, LightGBM, and XGBoost Regression. *Autom. Constr.* **2021**, *129*, 103827. [CrossRef]
40. Map. Available online: https://nr.indianrailways.gov.in/view_section.jsp?lang=0&id=0,1,285 (accessed on 22 January 2023).
41. Harrell, F.E. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*; Springer: New York, NY, USA, 2010; pp. 215–221.
42. Singh, G.; Pal, M.; Yadav, Y.; Singla, T. Deep Neural Network-Based Predictive Modelling of Road Accidents. *Neural Comput. Appl.* **2020**, *32*, 12417–12426. [CrossRef]
43. Ziegel, E.R.; Menard, S. Applied Logistic Regression Analysis. *Technometrics* **1996**, *38*, 192. [CrossRef]
44. Schumm, W.R.; Stevens, J. Applied Multivariate Statistics for the Social Sciences. *Am. Stat.* **1993**, *47*, 155. [CrossRef]

45. Hosmer, D.W., Jr.; Stanley, L. *Applied Logistic Regression*; John Wiley & Sons: New York, NY, USA, 2000.
46. Fawcett, T. An Introduction to ROC Analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [[CrossRef](#)]
47. Garson, G.D. Interpreting neural-network connection weights. *AI Expert* **1991**, *6*, 46–51.
48. Shahin, M.A.; Maier, H.R.; Jaksa, M.B. Predicting Settlement of Shallow Foundations Using Neural Networks. *J. Geotech. Geoenviron. Eng.* **2002**, *128*, 785–793. [[CrossRef](#)]
49. Jones, B.; Janssen, L.; Mannering, F. Analysis of the Frequency and Duration of Freeway Accidents in Seattle. *Accid. Anal. Prev.* **1991**, *23*, 239–255. [[CrossRef](#)]
50. Miaou, S.-P. The Relationship between Truck Accidents and Geometric Design of Road Sections: Poisson versus Negative Binomial Regressions. *Accid. Anal. Prev.* **1994**, *26*, 471–482. [[CrossRef](#)]
51. Pulugurtha, S.S.; Sambhara, V.R. Pedestrian Crash Estimation Models for Signalized Intersections. *Accid. Anal. Prev.* **2011**, *43*, 439–446. [[CrossRef](#)] [[PubMed](#)]
52. Chakraborty, A.; Mukherjee, D.; Mitra, S. Development of Pedestrian Crash Prediction Model for a Developing Country Using Artificial Neural Network. *Int. J. Inj. Control Saf. Promot.* **2019**, *26*, 283–293. [[CrossRef](#)] [[PubMed](#)]
53. Mukherjee, D.; Mitra, S. Impact of Road Infrastructure Land Use and Traffic Operational Characteristics on Pedestrian Fatality Risk: A Case Study of Kolkata, India. *Transp. Dev. Econ.* **2019**, *5*, 6. [[CrossRef](#)]
54. Priyadarshini, P.; Mitra, S. Investigating Pedestrian Risk Factors Leading to Pedestrian Fatalities in Kolkata City Roads. *Transp. Dev. Econ.* **2017**, *4*, 1. [[CrossRef](#)]
55. Soleimani, S.; Mousa, S.R.; Codjoe, J.; Leitner, M. A Comprehensive Railroad-Highway Grade Crossing Consolidation Model: A Machine Learning Approach. *Accid. Anal. Prev.* **2019**, *128*, 65–77. [[CrossRef](#)] [[PubMed](#)]
56. Soleimani, S.; Mohammadi, A.; Chen, J.; Leitner, M. Mining the highway-rail grade crossing crash data: A text mining approach. In Proceedings of the 18th IEEE International Conference on Machine Learning and Applications (ICMLA), Boca Raton, FL, USA, 16 December 2019; pp. 1063–1068.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.