

# Northumbria Research Link

Citation: Malik, Faheem, Dala, Laurent and Busawon, Krishna (2021) Real-Time Nanoscopic Rider Safety System for Smart and Green Mobility Based upon Varied Infrastructure Parameters. *Future Internet*, 14 (1). p. 9. ISSN 1999-5903

Published by: MDPI

URL: <https://doi.org/10.3390/fi14010009> <<https://doi.org/10.3390/fi14010009>>

This version was downloaded from Northumbria Research Link:  
<http://nrl.northumbria.ac.uk/id/eprint/48055/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)



## Article

# Real-Time Nanoscopic Rider Safety System for Smart and Green Mobility Based upon Varied Infrastructure Parameters

Faheem Ahmed Malik \*, Laurent Dala and Krishna Busawon

Northumbria Future Aerospace, Control and Transportation Research Group (n-FACT), Northumbria University, Newcastle upon Tyne NE1 8ST, UK; laurent.dala@northumbria.ac.uk (L.D.); Krishna.Busawon@northumbria.ac.uk (K.B.)

\* Correspondence: faheem.a.malik@northumbria.ac.uk

**Abstract:** To create a safe bicycle infrastructure system, this article develops an intelligent embedded learning system using a combination of deep neural networks. The learning system is used as a case study in the Northumbria region in England's northeast. It is made up of three components: (a) input data unit, (b) knowledge processing unit, and (c) output unit. It is demonstrated that various infrastructure characteristics influence bikers' safe interactions, which is used to estimate the riskiest age and gender rider groups. Two accurate prediction models are built, with a male accuracy of 88 per cent and a female accuracy of 95 per cent. The findings concluded that different infrastructures pose varying levels of risk to users of different ages and genders. Certain aspects of the infrastructure are hazardous to all bikers. However, the cyclist's characteristics determine the level of risk that any infrastructure feature presents. Following validation, the built learning system is interoperable under various scenarios, including current heterogeneous and future semi-autonomous and autonomous transportation systems. The results contribute towards understanding the risk variation of various infrastructure types. The study's findings will help to improve safety and lead to the construction of a sustainable integrated cycling transportation system.



**Citation:** Malik, F.A.; Dala, L.; Busawon, K. Real-Time Nanoscopic Rider Safety System for Smart and Green Mobility Based upon Varied Infrastructure Parameters. *Future Internet* **2022**, *14*, 9. <https://doi.org/10.3390/fi14010009>

Academic Editor: Mário Véstias

Received: 17 November 2021

Accepted: 22 December 2021

Published: 25 December 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 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/>).

**Keywords:** cyclist safety; road safety model; embedded learning system; infrastructure

## 1. Introduction

One of the significant challenges that the present world is facing is the uptake of sustainable modes of transport. The transportation sector is the second largest contributor to global greenhouse emissions [1]. Another challenge that needs immediate attention is the number of fatalities and casualties due to road traffic crashes. Road traffic crashes are the leading cause of death in the younger generation, 15–29 age group [2]. Hence, it is essential to improve safety as well as decrease the emissions from transportation. This will help save lives, decrease the number of injuries, and improve the cities' likeability. These objectives can only be achieved if we promote cycling as a mode of travel. The share of the cycling mode should increase many folds from the present low base [3]. Such a measure will have social, economic, as well as environmental benefits.

The primary hurdle that affects the uptake of cycling as a mode of travel is the high inequity of cyclists' safety. Cyclists face 13 times higher risk for the same distance traversed in Great Britain [4]. A study on the mode shift elasticity [5] concluded that the latter is more significant than that for cyclists, i.e., they attracted a proportionally higher number of cyclist road users for a safety improvement. Identifying the physical and environmental threats to the cyclist in the natural urban environment provides an insight into the preference and choice of cyclists [6]. The built environment, weather, work-related factors, and attitudes affect the everyday commute by bicycle [7]. In addition to the varying infrastructure, environmental, and traffic flow conditions, cycling hazards are also dependent upon cyclist-specific variables of age, experience, and gender [8]. It is widely established in the literature (see [9,10]) that how males and females use the infrastructure significantly varies, with varied safety implications.

Women are likely to make short journeys, and their journey's spatial and temporal structure is different from men. They rarely prefer large multi-lane roads and busy junctions. Instead, they prefer selected areas of the city having narrow streets with calming traffic measures. They generally cycle at lower speeds, are more likely to make recreational rather than commuter trips and have a stronger liking for quiet traffic streets [11]. A study [12] investigating crashes in the Czech Republic reported that males account for around 69% of the crashes and are more likely to be involved in a fatal crash (80%). Similarly, a study in the USA [13] found that males are at a higher risk than females (around five times more for the same distance travelled). It is common speculation that men drive less safely and more recklessly than women. Another reported variable in the literature is the age of the trip maker. The work on cyclist near-misses in London [14] led them to the conclusion that the age group of the rider directly affects their daily near-misses. They reported that the daily near-misses decrease with age, from 2.47 (20–29) to 1.85 (>60 groups). These near misses are correlated with crashes (see [15]). Similar results were obtained in Germany, and a study concluded that cyclists of varying age groups use the infrastructure differently and exhibit varying microscopic road traffic behaviour [16,17]. Men are more reluctant to modal shift to cycling than women [18,19], and it takes much more improvement in the infrastructure and environment for women to consider cycling [20]. The British cycling report (Transport Research Lab report 490) argued that gender, age, and cycling experience are critical variables affecting cyclists' safety. However, these variables do not influence how the cyclist rates a particular cyclable route. The qualitative evaluation of the infrastructure is the same across age and gender [21].

Cycling safety is an important topic; however, limited studies effectively model the risk in terms of exposure [22]. The present need of the transportation system requires cycling mode share to increase significantly. The major hurdle in this process is insufficient evidence to understand the relationship between cyclist safety and the identified parameters [23]. The literature is abundant with evidence of variables affecting cycling safety, which are presently not modelled in the prevalent general framework. The Italian crash study [24] demonstrated that the variable gender is a significant primary variable affecting the risk faced in terms of road type, type of interacting vehicle, riskiest vehicle manoeuvres, collision type, time, day, and season of the journey. The varied safety implication of gender is compounded by varied meteorological conditions that present different risks to different genders for varied infrastructure types [25]. Similarly, different lighting conditions also have different safety implications [26]. The needs of the cyclists are peculiar while interacting with other road users and infrastructure and its corresponding safety implications. The present road safety modelling framework implies a trade-off between prediction accuracy and knowledge of the causal link between the essential variables [27]. A practical model for examining road safety should determine causation, be highly predictive, and be scalable to large data sets [28]. This gap in the literature needs to be filled so that effective cycling infrastructure design and planning is not compromised. The transportation sector is at the brink of the fourth industrial revolution, with the movement toward autonomous/semi-autonomous vehicles and infrastructure to start soon. It is high time that the work on real-time embedded safety systems is explored. Presently, very few works in the literature undertake the development of an embedded learning system that can accurately model cyclist safety. Hence, the study aims to develop an intelligent embedded learning system for modelling cycling infrastructure. Such a system will be able to take the input data in real-time, model the variable, and result in an output of the predicted safety. As a result, the objectives are to:

1. Develop an intelligent framework to construct a real-time embedded learning system that accurately models cyclist safety.
2. Apply this learning system as a case study on an investigation area.
3. Construct a nanoscopic model for a cyclist to predict the safety for a particular age and gender.

4. Identify and quantify the significance of the variable affecting the unsafeness of the rider based upon the personal attribute.

The proposed embedded learning system will consist of specific hardware and software. This will result in a composite system that can continuously take the data and undertake modelling to present the policymakers/city planners with the final desired output that is ready to use. Such a proactive approach is necessary to achieve the 2030 vision of zero road traffic fatalities and embark on a pathway towards a smart, green, and integrated transportation system. It is high time that intelligent embedded systems are incorporated into transportation research as well as practice. The proposed real-time embedded learning system is described in the next section, followed by the results in Section 3, and the conclusions are drawn in Section 4.

## 2. Real-Time Intelligent Embedded Learning System

This section describes the proposed real-time intelligent embedded learning system in detail (Figure 1). It consists of three units: (a) input data unit, continuously collects data in real-time; (b) knowledge processing unit, develops an embedded learning system in real-time; and (c) output unit, results in the predictive model and variable interaction model.

### 2.1. Input Unit

The input learning system consists of automatic data collection unit, consisting of: (a) meteorological, (b) lighting, (c) traffic flow data, (d) crash data logger, and (e) digimap. The study area of north-east of England is used as a case study. Through the partnership with the city council, the metrological, lighting, crash datasets, traffic cameras, and counters were accessed. The flow characteristics for the study area were obtained from the traffic flow database system (TRADS). Through this system, the traffic cameras and counters were accessed (Figure 2). The exact WGS84 coordinates were extracted for each crash, which were then used as an input to obtain the concerned infrastructure parameters. Digimaps is an online map and data delivery service available to the research group. EDINA operates it at the University of Edinburgh. This platform was used to extract the infrastructure information based upon the WGS84 coordinates. This platform hosts accurate infrastructure maps depicting the present as well as past conditions. This ensured that exact infrastructure parameters were used for modelling based on the crash's temporal conditions rather than the present conditions. The sensors continuously feed the data into the KPU in the form of a combined base input file.

### 2.2. Knowledge Processing Unit

The first step in the knowledge processing unit (KPU) involves data cleaning. Multiple datasets are combined; hence an associated noise is present in the dataset, which is removed before data analysis has proceeded. The predictive models were developed through deep learning with neural network classifiers. The final input data was randomly divided into three categories of training, validation, and testing, in the ratio of 6.5:3.0:0.5. For the network to develop accurate prediction properties, this is the recommended division [29]. Such division guarantees adequate data for the network to learn correctly, evaluate the trained model, and apply the generated models to untrained circumstances. Bernoulli distribution was utilised to ensure that the data were divided randomly. Two different models (each for male and female) were constructed using the following input infrastructure variables (Table 1).

The output variables of the model are described in Table 2.

The following network structure was used to construct the model (Table 3).

The following four-step iterative learning process mapping input with the output variables was used.

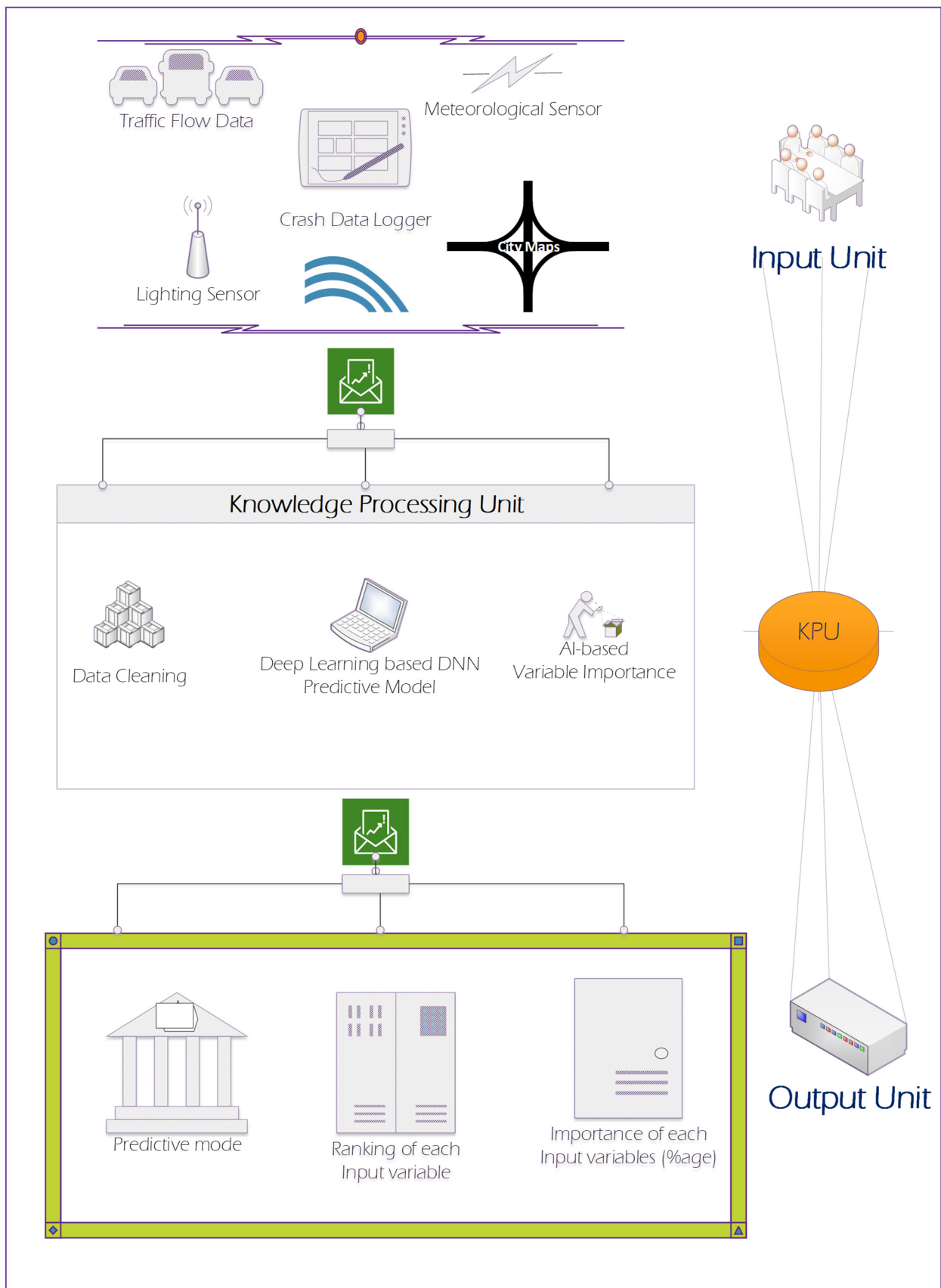


Figure 1. Real-time intelligent embedded learning system.

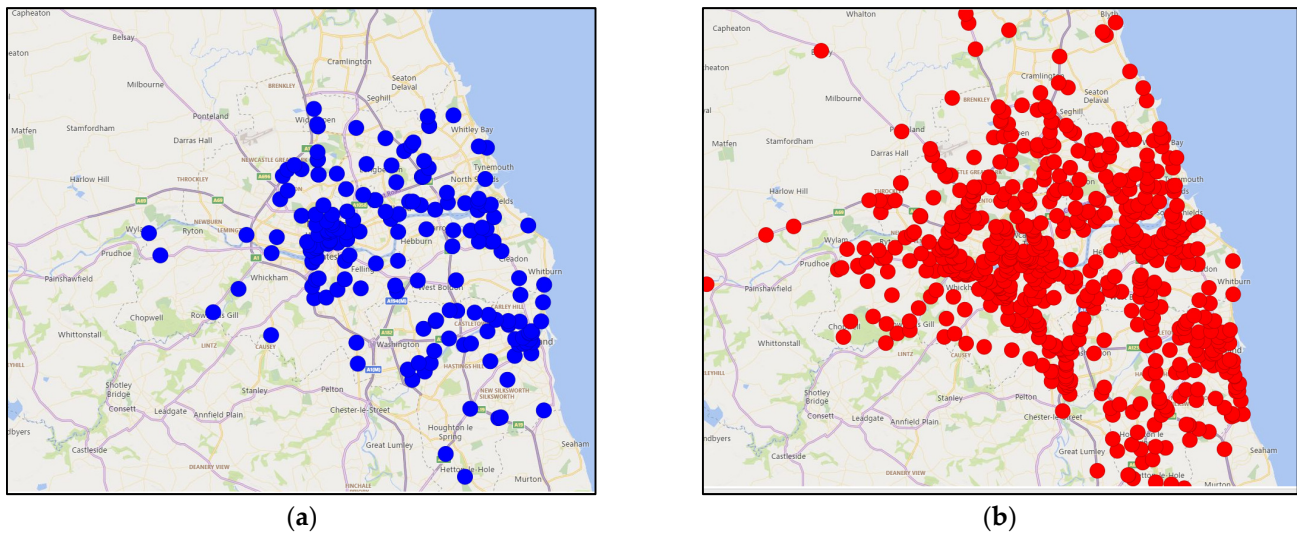


Figure 2. The overview of the (a) cyclist flow cameras and (b) motor vehicle cameras in the study area.

Table 1. Input variable for the proposed model.

No.	Input Variable	Values
(a)	Type of road	Dual carriageway, one-way street, roundabout, single carriageway, slip road
(b)	Speed limit	20–70
(c)	RC 1	A, B, C, E, U
(d)	$R_{HL}$	0–4
(e)	$R_{HLD}$	–4 to 4
(f)	Junction Detail	Crossroad, mini roundabout, multiple junction, straight road, roundabout, slip road, T or staggered, private drive
(g)	Junction Control	No control, traffic signal, give way or uncontrolled, stop sign
(h)	RC 2	A, B, C, E, U
(i)	Vehicle Manoeuvre	Changing lanes, going ahead, moving off, overtaking, parked, reversing, slowing/stopping, turning, U-turn, waiting to go ahead, waiting to turn
(j)	Vehicle Junction Location	Approaching junction or waiting/parked at junction exit, cleared junction or waiting/parked at junction exit, entering, leaving, mid junction, straight road (not at or within 20 m of the junction)
(k)	Road Location of vehicle	Bus lane, busway, cycle lane, cycleway, footpath, on layby or hard shoulder, main carriageway, tram/light rail track
(l)	Skidding and Overturning	No skidding or overturning or jack-knifing, overturned, skidded, overturned and skidded
<b>Output Variable</b>		<b>Risk gender and Age Group</b>

where RC = road class,  $R_{HL}$  = road hierarchy level, and  $R_{HLD}$  = road hierarchy level and direction change.

Table 2. Output variable.

Output Variable							
0–16 M	17–20 M	21–29 M	30–39 M	40–49 M	50–59 M	60–69 M	70+ M
0–16 F	17–20 F	21–29 F	30–39 F	40–49 F	50–59 F	60–69 F	70+ F

where M = male, and F = female.

**Table 3.** Deep learning network topography.

<b>Network Topology</b>	Number of hidden layers	2
	Elements in each layer	350
	Activation function between the hidden layers	Hyperbolic Tangent
	Activation function between hidden and output layer	Softmax
	Error function	Cross-entropy
<b>Stopping and Memory Criterion</b>	Steps (maximum) without a change in the error	999,999
	Training (maximum) time	999,999
	Training (maximum) epochs	999,999
	Relative change in the training error (minimum)	0.000001
	Relative change in the training error ratio (minimum)	0.000001
	Cases to store in the memory (maximum)	999,999
<b>Training</b>	Type	Batch
	Optimisation	Scaled conjugate gradient
	Initial Lambda	0.000000001
	Initial Sigma	0.000000001
	Initial Centre	0
	Initial offset	$\pm 0.000000001$
<b>Hidden layer (s)</b>	Total No. of hidden layers	2
	Total No. of units in the hidden layers	700 (350 in each layer)
<b>Output Layer</b>	Dependent variables	Age and Gender
	Total No. of output units	7

**Step 1:** Random weights and activation. Firstly, between each connection, i.e., input and hidden, within hidden layers, and hidden and output layer, random weights were assigned. For signal transmission between the synaptic cleft activation function ‘Hyperbolic tangent’ Equation (1) for hidden layers and ‘Softmax’ Equation (2) for the output layers were used.

$$F_{\epsilon} = \tanh(A_{\epsilon}) = \frac{e^{A_{\epsilon}} - e^{-A_{\epsilon}}}{e^{A_{\epsilon}} + e^{-A_{\epsilon}}} \tag{1}$$

$$\bar{F}_{\epsilon} = \sigma(\bar{A}_{\epsilon}) = \frac{e^{\bar{A}_{\epsilon}}}{\sum_{l=1}^j e^{\bar{A}_l}} \tag{2}$$

where  $F_{\epsilon}$  is the activation of the  $\epsilon$ th output neuron,  $j$  is the number of output neurons.

**Step 2:** Error modelling. Cross entropy error function was used to model the error between the output obtained and the desired output.

$$C = - \sum_{l=1}^j A_l \ln E_l \tag{3}$$

where  $E_l$  is the actual output obtained for the output node  $l$ , and  $A_l$  is the largest value of  $l$ .

**Step 3:** Synaptic weight update. The initially randomly assigned synaptic weights were updated based on the eq 3 error. The backpropagation algorithm calculates the gradient of the training error in each training case (epoch).

(i) nodes between the input and hidden layer

$$\frac{\partial C}{\partial s_{h\epsilon}} = \sum_{\epsilon=1}^m (F_{\epsilon} - t_l) x_h s_{h\epsilon} (1 - x_h) x_k \tag{4}$$

(ii) nodes between the output and hidden and layer

$$\frac{\partial C}{\partial s_{h\epsilon}} = (F_{\epsilon} - t_l)x_h \quad (5)$$

After error calculation, the weight ( $w_{h\epsilon}$ ) is updated in each epoch by adding it to the previously updated weight

$$\Delta s_{h\epsilon} = -\varnothing \frac{\partial C}{\partial s_{h\epsilon}} \quad (6)$$

$$\Delta s_{h\epsilon+1} = s_{h\epsilon} + \Delta s_{h\epsilon} \quad (7)$$

where  $\varnothing$  is the learning rate, and  $x$  is the input variable.

**Step 4:** Scaled conjugate gradient learning. The above steps were continuously repeated (iterated) until either the maximum number of these iterations (epochs) or minimum training error change was achieved.

The performance of the constructed predictive model was evaluated through the area under the curve (AUC) of the receiver operating characteristics (ROC). These are the evaluation matrices utilised for checking the networks classification performance. ROC is a probability curve, and AUROC represents the measure of the separability power of the network. While calculating the risk, the higher the AUROC value, the better is the distinguishable power of the network. Besides, gain and lift charts are used for qualitative evaluation; the visual aids for evaluating the performance assess the model's predictive capability compared with a non-model-based probability evaluation. After model construction and performance measurement, the next step is to validate the model through validation datasets. This process ensures an unbiased evaluation of the model fit on the training dataset while tuning the model hyperparameters, followed by checking the model's performance on unseen data, and providing an unbiased evaluation of the final developed predictive model.

The critical variables in the data learning model were identified through variable importance. The normalised significance of each variable concerning the most vital model was also evaluated to compare variables relative to each other. The independent variable importance is a measure of how much the predicted output value changes, viz a viz change in the input variable. The normalised significance of each input variable is their respective importance value divided by the largest importance value and expressed as percentages. This is followed by the Boolean logic, which presents the output in the form of the single most critical variable affecting the safety of a particular group.

The specific hardware for the embedded system was used. It has wireless connectivity with an inbuilt memory of 32 GB. Although the embedded system has its own processor, the said processor is not used in the KPU, which has its own separate processor and specific RAM. The use of the processor in the embedded system itself should be explored in future research.

### 3. Results and Discussion

There were 3325 bike collisions registered in the investigation area: 2638 being slight, 661 serious, and 26 fatal.

#### 3.1. Predictive Model

The findings of the deep learning prediction model that predicts the riskiest age and gender group based upon infrastructure input factors are reported in this section. Two deep learning models were built. Their accuracy is measured by comparing their distinguishable power between the riskiest and non-riskiest age and gender groups using AUROC, as shown in Table 4. The output predicted versus the observed for the testing datasets is presented in Figure 3, and the ROC curves for each output variable in Figure 4. It can be inferred from the ROC curves and the AUROC values that both the models attained a high level of accuracy. Each model can have up to seven different output values. As a result,



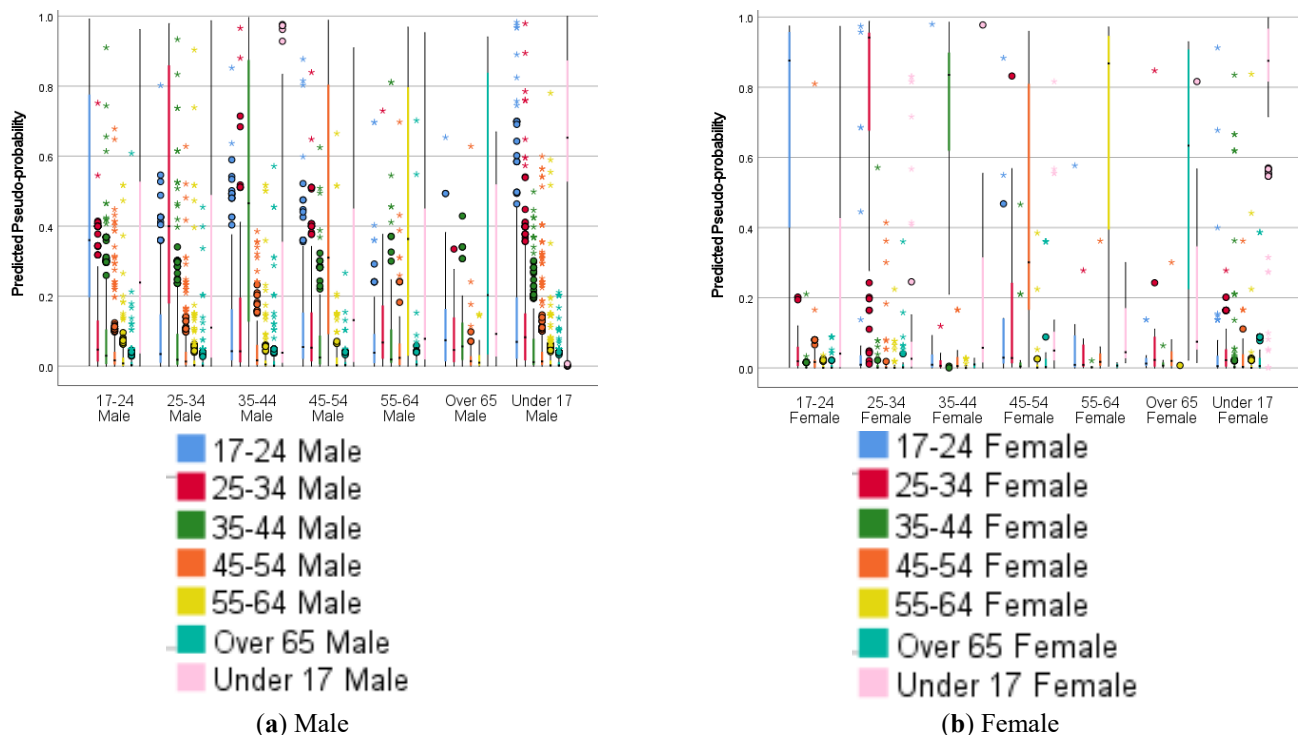
with perfect precision, the highest value of AUROC can be seven. The male and female AUROC values obtained were 6.19 and 6.64, respectively. This suggests that the prediction and distinguishable power is high and consistent. As the physical and cognitive abilities of riders of different ages and gender groups vary, there are safety implications for each rider sub-group. Hence, it makes it possible to estimate the most dangerous age and gender groups based on the individual input data.

The literature widely reports that current safety models cannot be directly utilised to simulate bike infrastructure, due to their inability to appropriately represent safety. In one comprehensive work to evaluate the efficacy of such a model, the Finnish crash model was gauged for its efficacy to model cyclist safety [30]. The study revealed an inaccuracy of around 65%. Similarly, the evaluation of the primary transportation simulation software, including PTV VISSIM, AIMSUN, TEXAS, and PARAMICS, exhibited an incapacity to mimic riders adequately and efficiently [31]. These studies concluded that cyclist interaction cannot be simulated using such packages, developed on the premise of motorised modes of transport. Hence, the accuracy achieved through the developed learning system is many times higher than the custom models in the literature. To further demonstrate the benefit of using the complex methodology for prediction, lift and gain charts are presented in Figures 3–6. These charts compare the efficacy of the output compared with the traditional probability-based statistical model.

**Table 4.** The AUROC values for each output variable of each model.

Male Prediction Model		Female Prediction Model	
Variable	AUROC	Variable	AUROC
Under 17 M	0.87	Under 17 F	0.94
17–24 M	0.87	17–24 F	0.94
25–34 M	0.87	25–34 F	0.95
35–44 M	0.90	35–44 F	0.96
45–54 M	0.89	45–54 F	0.93
55–64 M	0.86	55–64 F	0.96
Over 65 M	0.93	Over 65 F	0.96

where AUROC = area under the ROC curve, M = male, and F = female.



**Figure 3.** The predicted by observed for each output variable of: (a) male, (b) female.

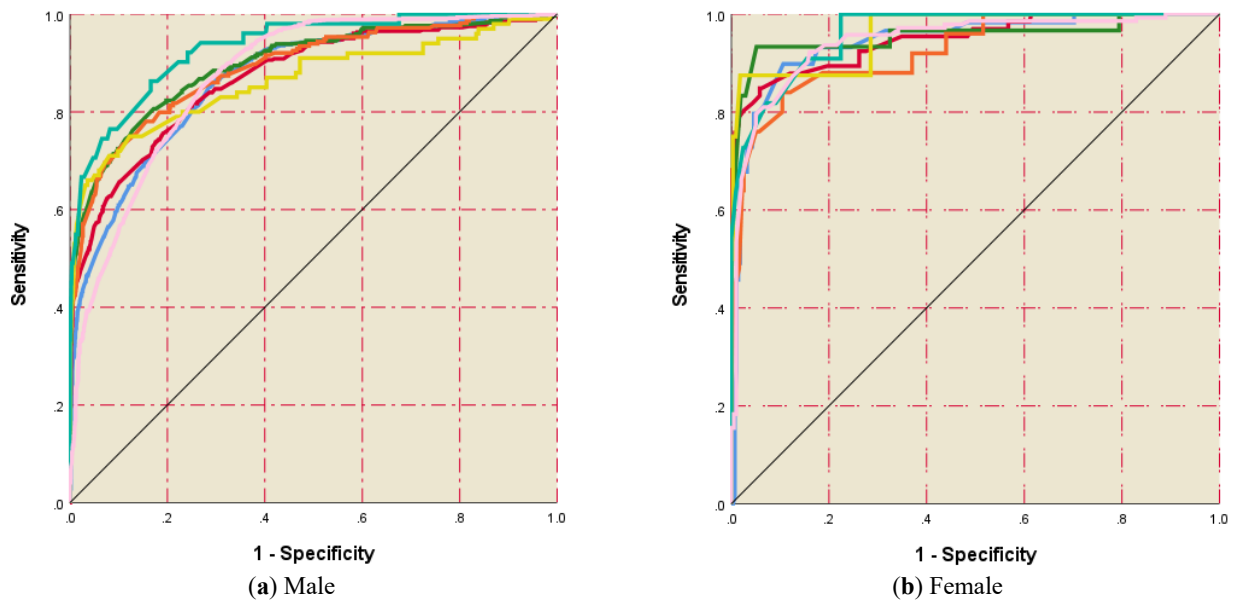


Figure 4. The ROC curve for each output variable of: (a) male, (b) female.

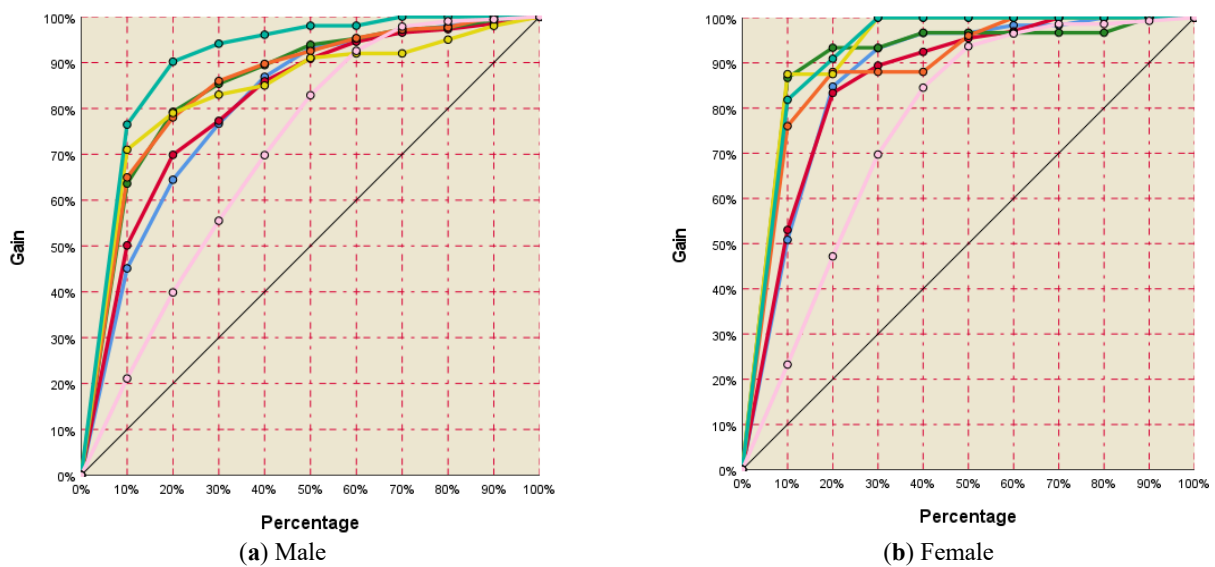


Figure 5. The gain chart for each output variable of: (a) male, (b) female.

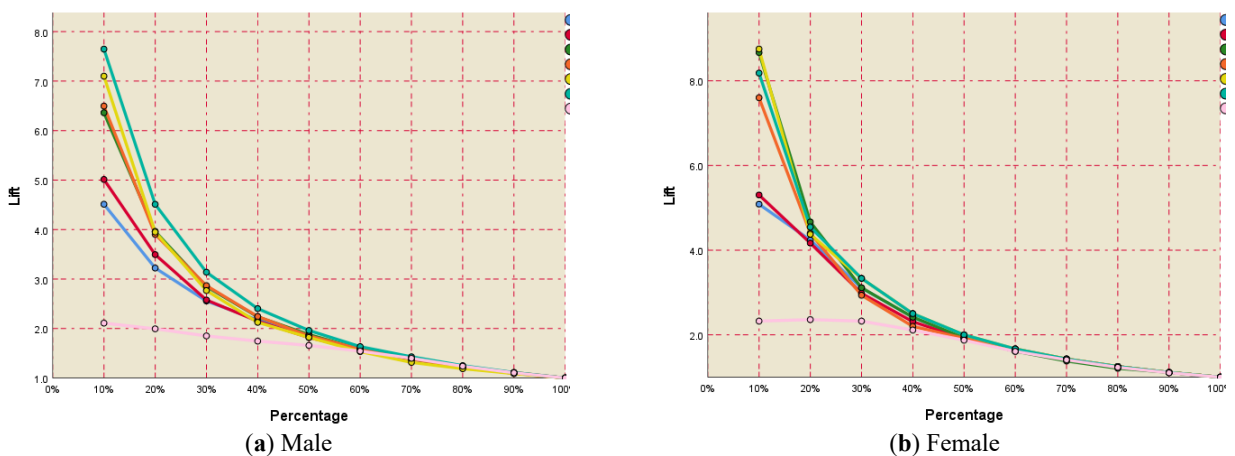


Figure 6. The lift charts for each output variable of: (a) male, (b) female.

For all the variable cases, the ROC curve line is closer to the upper left corner, farther away from the 45° baselines, which depicts significantly high prediction capability, evident from the AUC values. The gain is a measure of the effectiveness of the constructed model calculated as the percentage of the correct predictions obtained within the model versus the accurate predictions obtained without the model, i.e., baseline. A significant higher gain is obtained for all the output variables (an average of 10%, 80%), i.e., if we sort the wrong prediction by their pseudo probabilities, the top 10% of the dataset will have all 80% cases of improper predictions. Similarly, from the gain chart, the average gain value for the output, at 10% data is eight, i.e., the average accuracy of the models is eight times higher in comparison to the base case at this point.

### 3.2. Variable Interaction Model

The critical variables in the data learning model are identified through the variable and normalised significance (Table 5). This is based upon both testing and validation data sets. From the variable interaction model, it can be inferred that the overall effect of the infrastructure is not significantly different for riders of a different gender. However, the critical variables vary and differ by a small proportion in the rank of importance and their relative effect. The Boolean logic presents the most critical variable for males as  $R_{HLD}$ , and for females as vehicle manoeuvre. The following variables are estimated in the rank of their importance for males,  $R_{HLD}$ , followed by vehicle manoeuvre, junction location of the vehicle, junction detail, and road location of the vehicle. However, for females, critical variables are vehicle manoeuvre, followed by  $R_{HLD}$ , junction details, road location of the vehicle, and road type. The importance rank of the rest of the variables is similar; however, their normalised importance values vary slightly. This leads us to conclude that there are specific attributes of the infrastructure that are risky for all cyclists. However, the level of risk that each infrastructure attribute possesses is dependent upon the gender of the rider. The results agree with the findings in the literature. The infrastructural hazards present different levels of risk to the cyclist based upon their gender (see [20,32]), and that a bad infrastructure design/condition is rated poorly by the cyclists irrespective of their gender (see [23]). The novel variables introduced in the study, i.e., road hierarchy level and direction, are significant variables. It is recommended that these variables be considered in the cyclists' road safety investigations. The sudden change in the road hierarchy requires a shift in how cyclists need to interact with the infrastructure and other road users. Furthermore, the direction of change, i.e., whether the difference in the hierarchy is from a low class of road to a higher level or vice-versa, is a critical externality.

Table 5. The critical variable significance model for the input variable.

Male				Female			
R	Variable Importance (Descending Order)	V <sub>S</sub>	N <sub>S</sub>	R	Variable Importance (Descending Order)	V <sub>S</sub>	N <sub>S</sub>
1	$R_{HLD}$	0.098	100.0%	1	Vehicle manoeuvre	0.087	100.0%
2	Vehicle manoeuvre	0.096	98.2%	2	$R_{HLD}$	0.083	95.2%
3	Junction location of vehicle	0.094	96.8%	3	Junction detail	0.08	91.8%
4	Junction detail	0.092	94.5%	4	Road location of vehicle	0.072	82.3%
5	Road location of vehicle	0.086	88.4%	5	Road type	0.071	81.7%
6	RC 2	0.082	84.5%	6	RC 2	0.07	80.9%
7	Road type	0.081	83.3%	7	Speed limit	0.069	79.1%
8	Junction control	0.079	80.7%	8	Skidding and overturning	0.069	79.9%
9	Speed limit	0.076	77.6%	9	RC 1	0.067	77.2%
10	$R_{HL}$	0.075	77.1%	10	Junction location of vehicle	0.067	76.6%
11	RC 1	0.074	75.4%	11	Junction control	0.066	76.1%
12	Skidding and overturning	0.067	68.4%	12	$R_{HL}$	0.065	74.4%

where R = ranking, vs. = variable significance, and NS = normalised significance.

#### 4. Conclusions

This paper proposes a real-time nanoscopic rider safety system based on the varied infrastructure variables. Such a system will help improve safety, hence the attractiveness of the mode of travel. An embedded real-time learning system was developed, whose knowledge processing unit is primarily based upon deep neural networks. This is applied as a case study on the Northumbria region in England's northeast. Two accurate prediction safety models were developed that can predict the riskiest age and gender groups.

An average accuracy of 92% was obtained in these prediction models. This accuracy is several times higher than the available models in the literature. In addition, a variable interaction model was developed that ranks the input variable and estimates the variable importance.

The research's primary contribution is developing a comprehensive embedded system that models safety in real-time. The system includes an input unit, knowledge processing unit, and output unit. Not only prediction models are developed, but an understanding of the critical variables was determined and established. Presently, very few such systems exist in the literature that have affected infrastructure systems' design and planning. The present models in the literature are primarily based upon the probabilistic functions of human error, which model variables such as miles traversed, speed limits, intoxication, and others. These variables model road users' behaviour at an aggregate level, such as a city or a county. The learning system developed in this work models the road user at a nanoscopic level. The local authorities and city planners can directly apply such a system. Through the application of the learning system on the north-east of England, it is established that the varied types of infrastructure impose differing risks to riders of various ages. The interaction between cyclists and infrastructure can cause both physical and cognitive stressors, to which riders of different ages and genders respond differently, allowing us to estimate the riskiest age and gender group depending on the specific input factors. Furthermore, certain infrastructural characteristics are hazardous to all bicycles. However, the rider's personal attributes such as age and gender determine the amount of danger that any infrastructure poses. The critical variable affecting the safe usage of infrastructure for females is the rider manoeuvre that they are required to perform. This is due to the varying physical and cognitive abilities that females possess compared with males. However, for males, it is a sudden change in the road hierarchy level that presents the highest level of risk. The sudden change in the road hierarchy requires a shift in how cyclists need to interact with the infrastructure and other road users.

The study's findings will assist in improving safety and contribute to the development of a sustainable integrated bicycle transportation system. The study's findings will aid in better understanding the risk variance of various infrastructure types. Following validation, the developed real-time learning system is interoperable in several situations, including present heterogeneous and future semi-autonomous and autonomous transportation systems. The embedded learning system will improve the attractiveness of cycling as a mode of travel and contribute to a smart and green transportation system.

**Author Contributions:** Funding acquisition, L.D. and K.B.; investigation, F.A.M.; methodology, F.A.M.; supervision, L.D. and K.B.; writing—original draft, F.A.M., L.D. and K.B.; writing—review & editing, F.A.M., L.D. and K.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** The authors would like to thank Northumbria University for sponsoring the research through the research development fund and Gateshead City Council for accessing the TADU crash database.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** There is no conflict of interest to declare.

## References

1. Bell, M.; Sarkar, P.K.; Dilum, D.; Namdeo, A. A Methodological Framework to Estimate GHG from Travel Pattern of Tyne & Wear of Newcastle, UK with Various Policy Options using Fuzzy Logic Modal Split Model. *Int. J. Eng. Sci. Invent.* **2015**, *4*, 1–5.
2. IRF. *IRF World Statistics 2020*; IRF: Geneva, Switzerland, 2020.
3. Bell, M.C.; Galatioto, F.; Fryer, C.; Wang, R. The role of cycling in delivering sustainable travel by 2050. In Proceedings of the 48th Universities Transport Studies Group Annual Conference, Bristol, UK, 6–8 January 2016.
4. Department for Transport. *Reported Road Casualties in Great Britain, Provisional Estimates: Year Ending June 2019*; Department for Transport: London, UK, 2019.
5. Noland, R.B. Perceived risk and modal choice: Risk compensation in transportation systems. *Accid. Anal. Prev.* **1995**, *27*, 503–521. [[CrossRef](#)]
6. Lawson, A. *An Analysis of Cycling Safety and Development of a Bicycle Trip Assignment Methodology*; Trinity College Dublin: Dublin, Ireland, 2015.
7. Heinen, E.; Maat, K.; van Wee, B. Day-to-Day Choice to Commute or Not by Bicycle. *Transp. Res. Rec. J. Transp. Res. Board* **2011**, *2230*, 9–18. [[CrossRef](#)]
8. Bill, E.; Rowe, D.; Ferguson, N. Does experience affect perceived risk of cycling hazards? In Proceedings of the Scottish Transport Applications and Research (STAR) Conference, online, 15–17 December 2015; pp. 1–19.
9. Heinen, E.; van Wee, B.; Maat, K. Commuting by bicycle: An overview of the literature. *Transp. Rev.* **2010**, *30*, 59–96. [[CrossRef](#)]
10. Chen, H.; Chen, Q.; Chen, L.; Zhang, G. Analysis of risk factors affecting driver injury and crash injury with drivers under the influence of alcohol (DUI) and non-DUI. *Traffic Inj. Prev.* **2016**, *17*, 796–802. [[CrossRef](#)] [[PubMed](#)]
11. Beecham, R. *Exploring Gendered Cycling Behaviours within a Large, Attribute-Rich, Transactional Dataset*; UTSG: Edinburgh, UK, 2013.
12. Bíl, M.; Bílová, M.; Müller, I. Critical factors in fatal collisions of adult cyclists with automobiles. *Accid. Anal. Prev.* **2010**, *42*, 1632–1636. [[CrossRef](#)] [[PubMed](#)]
13. Rodgers, G.B. Bicyclist deaths and fatality risk patterns. *Accid. Anal. Prev.* **1995**, *27*, 215–223. [[CrossRef](#)]
14. Aldred, R.; Goodman, A. Predictors of the frequency and subjective experience of cycling near misses: Findings from the first two years of the UK Near Miss Project. *Accid. Anal. Prev.* **2018**, *110*, 161–170. [[CrossRef](#)] [[PubMed](#)]
15. Laureshyn, A.; Varhelyi, A. *The Swedish Traffic Conflict Technique: Observer's Manual*; Lund University: Lund, Sweden, 2018.
16. Schleinitz, K.; Petzoldt, T.; Franke-Bartholdt, L.; Krems, J.; Gehlert, T. The German Naturalistic Cycling Study—Comparing cycling speed of riders of different e-bikes and conventional bicycles. *Saf. Sci.* **2017**, *92*, 290–297. [[CrossRef](#)]
17. Potoglou, D.; Carlucci, F.; Cirà, A.; Restaino, M. Factors associated with urban non-fatal road-accident severity. *Int. J. Inj. Contr. Saf. Promot.* **2018**, *25*, 303–310. [[CrossRef](#)] [[PubMed](#)]
18. Welander, G.; Ekman, R.; Svanström, L.; Schelp, L.; Karlsson, A. Bicycle injuries in Western Sweden: A comparison between counties. *Accid. Anal. Prev.* **1999**, *31*, 13–19. [[CrossRef](#)]
19. Pedroso, F.E.; Angriman, F.; Bellows, A.L.; Taylor, K. Bicycle use and cyclist safety following boston's bicycle infrastructure expansion, 2009–2012. *Am. J. Public Health* **2016**, *106*, 2171–2177. [[CrossRef](#)] [[PubMed](#)]
20. Aldred, R.; Woodcock, J.; Goodman, A. Does More Cycling Mean More Diversity in Cycling? *Transp. Rev.* **2016**, *36*, 28–44. [[CrossRef](#)]
21. Guthrie, N.; Davies, D.G.; Gardner, G. *Cyclists' Assessments of Road and Traffic Conditions: The Development of a Cyclability Index*; TRL: Crowthorne, UK, 2001.
22. Aldred, R.; Goodman, A.; Gulliver, J.; Woodcock, J. Cycling injury risk in London: A case-control study exploring the impact of cycle volumes, motor vehicle volumes, and road characteristics including speed limits. *Accid. Anal. Prev.* **2018**, *117*, 75–84. [[CrossRef](#)] [[PubMed](#)]
23. TRL. *Infrastructure and Cyclist Safety: Research Findings*; TRL Report PPR 580; TRL: Crowthorne, UK, 2011.
24. Prati, G.; Fraboni, F.; de Angelis, M.; Pietrantoni, L. Gender differences in cyclists' crashes: An analysis of routinely recorded crash data. *Int. J. Inj. Control. Saf. Promot.* **2019**, *26*, 391–398. [[CrossRef](#)] [[PubMed](#)]
25. Pazdan, S. The impact of weather on bicycle risk exposure. *Arch. Transp.* **2020**, *56*, 89–105. [[CrossRef](#)]
26. Obeidat, M.S.; Khrais, S.K.; Bataineh, B.S.; Rababa, M.M. Impacts of roadway lighting on traffic crashes and safety in Jordan. *Int. J. Crashworthiness* **2020**, 1–10. [[CrossRef](#)]
27. Mannering, F.; Bhat, C.R.; Shankar, V.; Abdel-Aty, M. Analytic Methods in Accident Research Big data, traditional data and the tradeoffs between prediction and causality in highway-safety analysis. *Anal. Methods Accid. Res.* **2020**, *25*, 100113. [[CrossRef](#)]
28. Mannering, F.L.; Shankar, V.; Bhat, C.R. Analytic Methods in Accident Research Unobserved heterogeneity and the statistical analysis of highway accident data. *Anal. Methods Accid. Res.* **2016**, *11*, 1–16. [[CrossRef](#)]
29. Haykin, S. *Neural Networks, A Comprehensive Foundation*, 2nd ed.; Pearson Education: Singapore, 2005.
30. Peltola, H.; Kulmala, R. *Accident Models*; VTT: Espoo, Finland, 2010.
31. Gettman, D.; Pu, L.; Sayed, T.; Shelgy, S. *Surrogate Safety Assessment Model and Validation: Final Report, Report No. FHWA-HRT-08-051*; Research, Development, and Technology Turner-Fairbank Highway Research Center: McLean, VA, USA, 2008.
32. Abdel-Aty, M.A.; Radwan, A.E. Modeling traffic accident occurrence and involvement. *Accid. Anal. Prev.* **2000**, *32*, 633–642. [[CrossRef](#)]