Intelligent nanoscopic cyclist crash modelling for variable environmental conditions

Faheem Ahmed Malik
Mechanical and Construction Engineering
Northumbria University
Newcastle upon Tyne, UK
faheem.malik@northumbria.ac.uk

Laurent Dala
Mechanical and Construction Engineering
Northumbria University
Newcastle upon Tyne, UK
laurent.dala@northumbria.a.uk

Krishna Busawon
Mechanical and Construction Engineering
Northumbria University
Newcastle upon Tyne, UK
krishna.busawon@northumbria.ac.uk

***Abstract*— A cyclist is a vulnerable road user whose interaction with the road infrastructure depends on several factors, including variable environmental conditions of lighting and meteorological road surface. This paper is concerned with nanoscopic crash modelling under the riskiest environmental conditions. There are very few works in the literature dealing with such modelling. An intelligent methodological framework consisting of the data collection unit and a knowledge processing unit (KPU) is proposed. In the knowledge processing unit, a combination of a) statistical, b) data learning and c) casual inference methods are applied for investigating crashes on the study area of Tyne and Wear county in North-East of England. Three predictive nanoscopic road safety models are constructed (with 86% accuracy) using a) Spatial, b) Personal, and c) Infrastructure input variables. The importance of each of the identified input variable is estimated by deep learning and statistically validated through chi-square test and Cramer's statistic. It is found that unsafeness of interaction between rider and infrastructure depends on lighting and road surface meteorological conditions. Different environmental conditions present a varying degree of risk to different types of infrastructure. The riskiest environment conditions are significantly affected by rider's gender and age, traffic flow regime, specific riding manoeuvre, and the road hierarchy difference. The increase in the number of variables, a rider encounters during his entire trip, imparts risky riding behaviour, affecting its safe interaction with the infrastructure. A novel infrastructure variable, i.e. 'functional road hierarchy level and direction' introduced in this work , is found to be a critical road safety variable. A shift in road safety analysis towards nanoscopic modelling can help achieve zero-vision road traffic fatality. The study reinforces the need to plan and design infrastructure to move towards a more holistic approach while considering this vulnerable road user's limitations.**

Keywords*— intelligent transportation system, cycling safety, nanoscopic safety modelling, environmental conditions

I. Introduction

Creating a complete and comprehensive network for cycle traffic is imperative, which is both comfortable and attractive for the user [1]. Based on its role in providing a sustainable transportation system, bicycling has started gaining a prominent transportation policy role. To embark on a pathway towards a sustainable transportation system, cycling mode share has to increase by many folds [2], which will reduce carbon footprint and enhance cities' liveabilities. However, safety concerns are associated with this mode of transportation, which is the most commonly perceived barrier to its uptake , [2], [3]. The road traffic crashes have adverse effects on human health, the wellbeing of individuals, and society. Crashes have associated pain, grief and suffering due to personal injuries, property damage, increased travel time, and a corresponding increase in carbon emissions due to congestion. Road safety involves a complex interaction of various factors and underlying phenomena, requiring an in-depth understanding and knowledge-driven measures to reduce crashes' frequency and impact. The preference and requirements of cyclists are different from other road users [4]. Safety is also a significant mode and route choice variable [5]. The effect of safety pessimisticism is a more considerable deterrent than the effort involved in riding [6]. The susceptibility of the cyclists towards different externalities is more pronounced than the motorists (see [7]–[9]). These externalities include different infrastructure types, personal attributes of the rider, traffic flow regime, and, variable weather and lighting conditions. Very few works have either reported or attempted to model the former variable of varying environmental conditions.

The literature has widely reported that extended periods of rainfall negatively affect cycling, affecting the selection of cycling as a mode of travel and its safe usage of infrastructure [10]. The English and Wales mode choice model [11] reported that rainfall has high negative cyclist flow elasticity. The variable environment conditions can result in an additional variable for the cyclist to deal/ negotiate with while interacting with the infrastructure under different traffic flow regimes; thereby acting as a significant hazard. This phenomenon can be attributed to the safety law of complexity [12]; 'more the variables road user has to attend to; notable is the risk faced. The rain degrades the driving environment through various physical factors, through a possible loss of friction between the tyre and road, impaired visibility, and a spray of water from other vehicles [13]. These conditions can also impact the cyclists riding comfort [14] , its cognitive capability (safety law of cognitive capacity), making it a potential safety hotspot. These can affect the safety variedly for a cyclist varying from one rider to another [15].

The study in Finland on road traffic crashes from 2014 to 2016, to assess the impact of road surface conditions on road safety demonstrated that crash risk increases due to poor road weather conditions. They reported that the risk of adverse environmental conditions on different infrastructure is highly varied in crash frequency and impact [16]. The longitudinal survey in the Netherland [15] investigating a cyclists day to day choices found that the cyclist, especially women riders, are significantly affected by the absence of daylight, affecting their perceived safety and corresponding mode choice. Another similar study found that extreme weather conditions substantially increase the crash rates (20% risk increase). The effect of variable environmental conditions on safety varies spatially with month and day of the week journey is being undertaken [17]. An analysis of the impact of road surface condition on road safety in Iowa (USA), found pavement skid resistance has a significant impact on crashes under varied environmental conditions [18].

This combination of different factors can lead to increased crash rates due to a strain on the road user's cognitive capability [12]. Different road users respond differently to these road safety variables, which is also evident in their route choices [19]. Hyden's safety pyramid [20] and the Swedish traffic conflict technique [21] demonstrate a pyramid-shaped relationship between the crashes and everyday conflicts that the cyclist faces with other road users. These conflicts form the pyramid's base, potential crashes at the top, whereas crashes are the tip; both these variables are interlinked and causatives. It is well established from the literature that the cyclists' conflicts vary from a user to user depending upon their personal attributes [22], [23], traffic flow regimes [24], and type of infrastructure [25]. Therefore, cyclists' safety modelling needs to be incorporate the same and thus, be more comprehensive with the modelling focus on a cyclist at the nanoscopic level, rather than the aggregate usage of infrastructure at the macro level. The increased risk reported in the literature due to varied environmental conditions of lighting and meteorological road surface conditions is summarized in Table I.

Table I. Increased risk due to varied environment conditions (lighting and meteorological road surface condition)

Study location	Period	Relative Crash Rates	Citation
West Virginia, USA	1970	2.2	[26]
Glasgow, UK	1978-79	1.2-1.3	[27]
Chicago, USA	1977-79	2.0	[28]
Edmonton, Canada	1983	1.3-1.9	[29]
Canada	1995-1998	1.75	[30]
Melbourne Australia	1987-2002	1.61-1.67	[31]
Iowa, USA	1965-2005	1.84	[32]
Vancouver, Canada	2003-2007	1.13-1.55	[33]
Finland	2000-2010	1.20	[17]
New Zealand	2012	1.35	[34]
China	2001-2016	1.13	[35]
Jordan	2020	2	[36]

Therefore, after establishing the importance of the variable environmental conditions on cycling safety, and need to perform modelling at the nanoscopic level, it is imperative to understand the modelling procedure. The infrastructure safety analysis is performed by developing safety models, whose accuracy and efficiency directly impact road safety investigations, remedial measures, planning, and design. The first mathematical theory used in safety modelling is generalized linear modelling. Over time, various studies proposed a generalized linear model assuming a non-normal error structure [37], overcoming linear regression limitations and produced a better fit to the observed collision data [38]. As crashes are discrete positive integral variables, therefore this prompted the use of Poisson regression. However, it is unable to handle overdispersion (i.e. the variance exceeding the mean). This then motivated using negative binomial or Poisson gamma models, assuming that Poisson parameters follow a gamma distribution [39]. However, there are profuse locations with zero reported crashes, which motivated the zero-inflated Poisson method, having two different states; zero state and normal count state. For improving the modelling capabilities, various techniques such as hierarchical, random effect, cart, finite-mixture/ latent-class, log-linear, probit/logit, Markov switching, Poisson–Log normal Regression, Empirical Bayes Method, Conway-Maxwell-Poisson, negative Binomial-Lindley method and others [40], [41], have been explored in the literature. However, as cycling safety is a multifactor phenomenon based upon the complex non-linear interaction of variables, very few techniques are able to model them accurately.

This limitation is quite evident in the present crash models, which are mostly reactive [42] and unable to consider the cyclist's dynamic interaction nature with variable infrastructure and quantify its safety implications [43]. These conventional models are generally developed for assignment of motorized modes of travel and are not equipped for a cyclist's needs [44]. Although there have been attempts to the model cyclist's safety, however, from the literature, it is evident that the cyclist's specification and requirement are different from motorists [45]. For a modeller to effectively and efficiently model cyclist safety, the cyclists' vulnerability and susceptibility to various externalities need to be modelled at nanoscopic level. One of the established externality is the variable environmental conditions for a cyclist. Presently, there are limited studies which have or attempted to undertake such modelling.

Hence, we can establish from the literature; there is an *"Absence of an accurate and dynamic model for a cyclists, which can model varied environmental conditions"*. This study seeks to improve the understanding of how environmental conditions influence particular cyclists' road safety. Therefore, the aim is to develop a road safety model for a cyclist at the individual level (nanoscopic), predicting the environmental conditions most likely to be unsafe, based upon the specific input variables and investigating the causal relationship between the input variables and riskiest environment condition. More precisely, our objectives are:

- Investigate, and develop an understanding of how cyclist's safety varies with varying environmental conditions.
- Test the hypothesis that unsafeness of interaction between rider and infrastructure depends on the lighting and meteorological road surface condition.
- Develop a nanoscopic road safety model with environment condition as an output variable.
- Identify the most critical variable affecting the unsafeness during a prevalent environmental condition for an individual.
- Validate the results statistically.

The aim requires developing a model, which can only be performed through an application on a study area. The study area of the northeast of England is selected for investigation. This is one of the nine official regions of England: population of 1.13 million and 3,317 sq. miles area. There are five boroughs in the study area, having thirteen urban, and three rural districts (Fig 1). In the next section, the proposed intelligent hybrid safety modelling framework is de

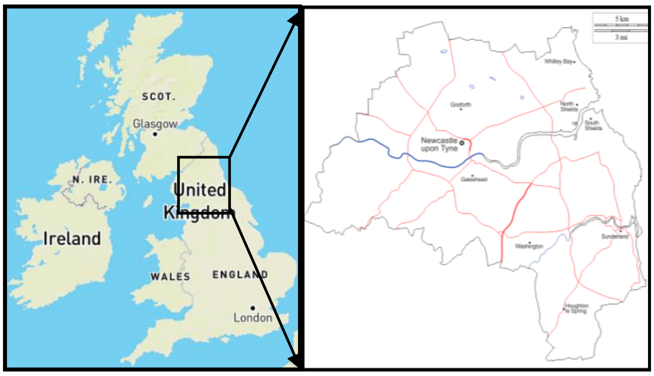


Fig. 1. Study area for model development; northeast of England.

-fined on the study area. The results and discussion are described in section III, limitations in section IV, and conclusions drawn from the study in section V.

II. INTELLIGENT HYBRID MODELLING FRAMEWORK.

The proposed intelligent hybrid modelling framework consists of a) Data collection, and b) Knowledge processing unit (KPU.)The study area's crash dataset from 2005 to 2018 is used for modelling. The data concerning each crash is collected through an in-depth crash investigation of the site, performed by the city council and Northumbria police force. The attributes of each crash are recorded in a predefined document 'STATS 19'. For each crash, following four types of information is recorded: i) Type of severity, ii) Time, date, and location of the crash, iii) Environment conditions such as lighting, weather, road surface condition, type of infrastructure and number of vehicles involved, iv) Sociodemographic information such as age, gender, intoxication, journey purpose of the cyclist. All the STATS 19 forms are stored on an online platform, housed by Department for Transport (DfT). For this study, we are provided access to the crash database Traffic and Data Unit (TADU) by the city council. The classification of the severity is performed through the Department for Transport (DfT) criterion; fatal: if crash results in the death within 28 days of the crash, serious: crash resulting in death either after 28 days or an overnight admission in the hospital, and slight: crash resulting in overnight hospital discharge or property damage only [46]. The crash investigation by DfT aims to record the information as accurately as possible, as it serves the basis for further legal and other courses of actions

To obtain detailed information concerning the crash site's infrastructure, WGS84 coordinates of each crash are extracted from TADU. These GPS coordinates are recorded as accurately as possible, as it serves the basis for further legal and other courses of actions. These coordinates are used to obtain detailed infrastructure parameters using Digimap (access obtained for the study), an online platform housing the UK roads' macro and micro characteristics including the control, and public transport details. This platform also houses historical and spatial data. For each crash, requisite infrastructure parameters are extracted for the concerned crash date, and then correlated with the maintenance plan (from city council) to check any maintenance work that might have affected the infrastructure's usage. This ensures that the crash time's infrastructure parameters are modelled rather than the current features, which may have varied over time. Finally, all the crash details are combined in a base file to be used as an input file in the Knowledge Processing Unit (Table II).

Knowledge Processing Unit,

There are four general methodological frameworks for road safety analysis i) Traditional statistical, ii) Heterogenetic modelling, iii) Causal inference and iv) Data-driven framework. The process of selecting one of these involve making an implicit tradeoff between the prediction accuracy and understanding of the causal relationship between the governing variables [47]. The data-driven methods have a proven application in engineering due to their ability to handle a large amount of data with high prediction accuracy. However, these are a 'black box', due to their inability to understand variable interaction and contribution of each variable. On the other hand, the casual-inference framework is better able to identify and explain the underlying phenomenon. Yet, these have been rarely used in safety modelling, due to weak predictive capability, and ability to address limited explanatory variables. Therefore, a hybrid methodology is proposed consisting: i) Traditional statistical, ii) Data-driven, and iii) Causal inference frameworks. This hybrid methodology can investigate the causal relationship between governing variables, have a high predictive capability, and scalable to a large data set

A. Traditional Statistical Framework

Firstly, statistical analysis of the crashes is undertaken, followed by the generation of the heat maps. This results in crash rates and investigates the risk's spatial variation for varied environmental conditions with different infrastructure.

B. Data-driven (Deep Learning Method)

A predictive model is developed using supervised deep learning neural network classifier, and gradient descent backpropagation error function. It is the sub-group of the machine learning techniques based upon computational methodologies imitating the human brain's working. These are massive parallel distributed processors that have a natural propensity to store experiential knowledge [48]. The road safety problem is highly non-linear and characterized by the underlying correlation between various infrastructural, environmental, and personal attributes of the rider. Deep learning can capture this non-linear and complex underlying characteristic, with a high level of accuracy [49]. The neural network has been widely applied as a data analytic method in transportation science [50], resulting in generic, accurate, and convenient mathematical models, simulating the numerical model components [51]. This is due to their ability to work with a large amount of the multi-dimensional data, modelling flexibility, learning, generalization ability, adaptability and good predictive capacity [51].

In the first step of model development, a learning algorithm is developed to divide the data set randomly into training (65%), validation (30%), and testing (5%). The division ensures proper learning of the constructed model, assessment of the trained model, and ensures that the constructed models are relevant to the untrained scenarios [50]. Three models are built using a) Spatial, b) Personal, and c) Infrastructure dynamic input variables, described in Table II.

Table II. Input variable for the proposed model

No.	Input Variable	Values
1.	**Spatial**	
1.1	Journey Hour (The hour in which the crash has occurred)	0-23.
1.2	Number of vehicles (Number of vehicles involved in the crash).	1-5.
1.3	Month of Journey (Month in which the crash has occurred).	Jan-Dec.
1.4	Journey Day (Day of the week on which crash has occurred. The day, month and hour of the journey are a representation of the traffic flow regime that was plying at the time of the crash)	Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday.
1.5	Journey Weekday/ Weekend	Weekday. Weekend.
2.	**Personal**	
2.1	Gender (Gender of the rider).	Male, Female, and Unknown
2.2	Age (Age of the rider)	0-17, 18-24,25-34, 35-44, 45-54,55-64, and over 65.
2.3	Age and Gender (Combined)	0-17 male, 14-24 male,25-34 male, 35-44 male, 45-54 male,55-64 male, over 65 male, 0-17 female, 14-24 female,25-34 female, 35-44 female, 45-54 female,55-64 female, and over 65 females.
2.4	Journey Purpose (The purpose of the journey being undertaken in which the crash has occurred).	Commuting, work trip, School Journey by Pupil, taking pupil to school, other, Unknown.
3.	**Infrastructure**	
3.1.	Road Type (Type of road infrastructure present at the crash spot).	Dual Carriageway, One-way street, Roundabout, single carriageway, slip road,
3.2.	Speed limit (maximum permissible speed limit on the road).	20-70
3.3.	1st Road Class*	Functional classification of the roadway into: A, B, C, E, U
3.4.	Road Hierarchy Level* (The difference in the functional classification of 1st and 2nd road class)	0-4
3.5.	Road Hierarchy level and direction*(The difference in the functional classification of 1st and 2nd road class including the direction of change)	-4 to 4
3.6.	Junction Detail (Type of intersection).	Crossroad, Mini Roundabout, Multiple Junction, Straight Road, Roundabout, Slip Road, T or Staggered, Private Drive
3.7.	Junction Control (Type of control that is employed at the intersection).	No Control, Traffic Signal, Give way or uncontrolled, Stop sign
3.8.	2nd Road Class*.	Functional classification of the roadway into: A, B, C, E, U
3.9.	Vehicle Maneuver (The Maneuver that the rider was performing/intending to perform when the crash occurred).	Changing lanes, Going ahead, Moving off, Overtaking, Parked, Reversing, Slowing/stopping, Turning, U-turn, Waiting to go ahead, waiting to turn
3.10.	Vehicle Junction Location (Location of the cyclist to the junction when the crash has occurred).	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 meters of the junction)
3.11	Road Location of vehicle (Location of the cyclist to the road infrastructure, when the crash has occurred).	Bus Lane, Busway, Cycle lane, cycleway, footpath, on layby or hard shoulder, main carriageway, tram/light rail track
3.12.	Carriageway Hazard (Additional unexpected hazards on the carriageway).	Animal in the carriageway (except ridden horse), Dislodged vehicle load in carriageway, Dislodged vehicle load in carriageway, Other object in carriageway, and none
	Output Variable	**Riskiest Environment conditions**

An intersection is a roadway facility, wherein two or more roads either meet or intersect each other. These intersections can have two different hierarchies of the road network which join or intersect. The *variable first road class* is the functional classification of the roadway on which the vehicle was travelling when the crash occurred. The *second road class* is the functional classification of the roadway on which the vehicle intended to move to (if the crash happened before exiting the intersection) or roadway on which cyclist came from (if the crash occurred after negotiating the intersection). The cyclist while utilizing the infrastructure are frequently required to change the type of road using, e.g., rider may change from an A road type (which is main collector road between the cities) to a one-way estate street (E road). This sudden transformation can result in a change in the cyclist's behaviour and other road users' interaction behaviour with the cyclist. This transformation needs to be modelled effectively, as it has the potential to impact safe interaction negatively. Therefore, we are introducing a new variable *road hierarchy level*, for the difference in the road hierarchy between the first and second road class. This road hierarchy level only signifies the change in the functional hierarchy, e.g., if the road user is moving from A-type of the road to E class of road, is modelled in the same manner as that of moving from E class to A-type. This motivates introducing an advanced form of this variable, i.e., *the road hierarchy level and direction*. It also considers the direction of the change in the road hierarchy, i.e., whether a cyclist is moving to a higher road class or a lower road class. If the change in the functional hierarchy is one level, and it moves from higher to the lower class, then it is modelled as -1, whereas if it moves to one higher level, it is modelled as +1. Similarly, all crashes are modelled from -4 to 4.

The primary objective for developing the predictive environmental condition model is to aid in in the inclusion of this variable in the development and design of cycling infrastructure. We can then further identify the underlying safety performance functions, if we can demonstrate that this variable can be modelled effectively and efficiently. The input variables can be adapted in inverse analysis to model reduction in the riskiness of a particular infrastructure during severe environmental circumstances and develop recommendation improvements and management strategies. The cyclists follow Swiss cheese safety criterion; therefore, any minor change in the input variables can decrease the overall risk. Modelling environmental conditions is a first step towards developing a real-time safety model. In such a model, the safety will be predicted in real time, leading to a nanoscopic real time route selection for a particular rider (subject of a further paper). The model has a renewed focus; as we progress toward autonomous vehicles and infrastructure management. The motorist-cyclist algorithms can be developed to account for additional safety margins under these scenarios (predicted by the predictive model). Such a nanoscopic approach is critical if we are to achieve the vision of zero road traffic fatalities. The safety, mode and route choices are correlated; therefore, the model will aid in better flow modelling and better management of the network. Table III shows the output variables that the predicted model can take

Table III. Risky Environment Conditions (Light and Meteorological Road Surface Condition)

Output Variable: Riskiest Environmental condition of	
Darkness - No Street Lighting, and Dry	Darkness - Street Lights present, unlit and Dry
Darkness - No Street Lighting, and Wet/Damp	Darkness - Street Lights present, unlit and Wet/Damp
Darkness - Street Lighting Unknown, and Dry	Daylight and Dry
Darkness - Street Lighting Unknown, and Wet/Damp	Daylight and Frost
Darkness - Street Lights present, lit and Dry	Daylight and Snow
Darkness - Street Lights present, lit and Snow	Daylight and Wet/Damp
Darkness - Street Lights present, lit and Wet/Damp	n/a

The neural network consists of neurons grouped into different interconnected layers of input, hidden and output layers. The neurons from one layer interact with neurons from other layers through weighted connection, a real number signifying the strength of association and its relationship. A neuron from a single layer is attached to multiple neurons from the previous layer. In this manner, the signal flows throughout the network. Through these weighted connections, the networks learn to map the given input with the output and perform non-linear mapping of a higher differential order, which cannot be undertaken using simple conventional mathematical theories. The following four steps iterative process is used for modelling the input (Table II), with the output (Table III) variables

Step 1: Firstly, random weights are assigned to each connection between the input and hidden, first and second hidden, and second hidden and output layer. For signal transmission within the layer's activation function, 'Hyperbolic tangent' for the hidden layers (eq 1), and 'Softmax' for the output layer (eq 2) is used.

$$A_j = \tanh(S_j) = \frac{e^{S_j} - e^{-S_j}}{e^{S_j} + e^{-S_j}} \quad (1)$$

$$A_j = \sigma(S_j) = \frac{e^{S_j}}{\sum_{k=1}^{m} e^{S_k}} \quad (2)$$

where A_i is the activation of the ath output neuron, and m is the number of output neurons. These functions take real numbers as arguments and return real values [-1, +1].

Step 2: Cross-entropy error function (eq 3) is used to obtain the error (as initially weights are randomly assigned) between the desired output (target) and output achieved.

$$E = -\sum_{j=1}^{m} t_j \ln O_a \quad (3)$$

where O_a is the actual output obtained for the output node j, t_j is the largest value of j, and m is the number of nodes.

Step 3: The initial randomly assigned weights are updated based upon the error achieved in step 2. The backpropagation algorithm is used to determine the training error's gradient in each training case (epoch).

a) nodes between the input and hidden layer

$$\frac{\partial E}{\partial w_{hj}} = \sum_{j=1}^{m} (O_a - t_j)\, x_h w_{hj} (1 - x_h) \quad (4)$$

b) nodes between the output and hidden layer

$$\frac{\partial E}{\partial w_{hj}} = (O_a - t_j) x_h \quad (5)$$

In each of the training case (epoch), The weight w_{ih} is updated continuously in each epoch, by adding it

$$\Delta w_{ih} = -\gamma \frac{\partial E}{\partial w_{hj}} \quad (6)$$

$$\Delta w_{ih+1} = w_{ih} + \Delta w_{ih} \quad (7)$$

γ is the learning rate, and x is the input variable.

Step 4: The weights are continuously updated, i.e., iteration is performed until either minimum change in training error or the maximum number of iterations (epochs) condition is fulfilled.

The network structure to construct the model is explicitly defined in Table IV.

Table IV. The network structure of the deep learning model

Network Topology	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
Training	Type	Batch
	Optimization	Scaled conjugate gradient
	Initial Lambda	0.000000001
	Initial Sigma	0.000000001
	Initial Centre	0
	Initial offset	±0.000000001
Stopping and Memory Criterion	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

The recommended methodology for measuring the performance of the constructed model is to develop Receiver Operating Characteristics (ROC) curve [52], which gives the visual display of sensitivity and specificity.. Sensitivity is a measure of true positive cases that are predicted by the model as positive, whereas specificity is the measure of the true negative cases. To evaluate the performance quantitatively, Area Under the Curve of the ROC (AUROC) is used, an evaluation matrix utilized for checking networks' classification performance. ROC is a probability curve, and AUROC represents the measure of the separability power of the network. In calculating the risk, the better the AUROC value (closer to 1 (100%)), the better is the network's distinguishable power between the risky and non-risky environmental condition. Besides, gain and lift charts are used for qualitative evaluation, the visual aids for evaluating performance. After constructing the model and measuring the performance, the next step is to validate the model through validation datasets. This process ensures an unbiased evaluation of the fitted model on the training dataset while tuning the model hyperparameters. This is followed by checking the model's performance on unseen data, providing an unbiased evaluation of the final model constructed using the training dataset. Through this three-step process of training, validation, and testing, the constructed model's performance is estimated to establish the credibility and confidence for further evaluation, planning, design, and policy implications.

C. *Causal Inference*

The critical variables in the data learning model are identified through variable importance, and normalized importance of each variable concerning the most critical variable is also calculated. This is based upon both testing and validation data sets. The independent variable importance is a measure of how much the predicted output value changes viz a viz change in the input variable. The normalized importance of each input variable is their respective importance value divided by the largest importance value, expressed as a percentage.

These critical variables identified, need to be validated statistically for their association. These are measured either on a nominal or ordinal scale. Therefore, non-parametric technique is the ideal statistical method in such a case, especially when the sample size is small. However, two assumptions need to be met i) Random sampling, and ii) Independence of observations. The crashes are a random phenomenon [53] and are independent of other crashes occurring at different locations, thereby, satisfying the two requisite requirements. The chi-square test for goodness of fit is a non-parametric technique, which tests whether there exists a relationship between the two variables and uses the sample data to test the hypothesis regarding the shape of the proportion of population distribution. It determines how well the obtained sample proportions fit the population proportion specified by the null hypothesis. Each variable in the sample is classified on n variables, creating an n-dimensional frequency distribution matrix. Whenever chi-square test involves a matrix larger than two by two order, modification of the Phi-Coefficient known as Cramer's V, is used to measure the strength of association [54]. The following four-step statistical approach is used:

Step 1: Degree of freedom of the two variables, whose association is being evaluated is calculated:

$$df = (R-1)\ V\ (C-1) \tag{8}$$

where R = row, and C = column

Step 2: Chi-square statistic is calculated:

$$\chi^2 = \sum \frac{\left(n_{ij} - \frac{n_i n_j}{n}\right)^2}{\frac{n_i n_j}{n}} = \sum \frac{(Observed - Expected)^2}{Expected} \tag{9}$$

Step 3: For determining the strength of the correlation, Cramer's V statistic is calculated:

$$V = \sqrt{\frac{\chi^2}{n(df)}} \tag{10}$$

Step 4: Strength of correlation: Cramer's V is a single-valued output, which is converted into qualitative knowledge through Cohen's table. This determines the strength of correlation using the degree of freedom and numerical *V* value, in terms of no correlation, small, medium, and large correlation.

III. RESULTS AND DISCUSSION

A. *Statistical Model*

There are 3,325 (79.3% slight, 19.9% serious, and 0.8% fatal) cyclist crashes reported in the study area. Out of these, 83 % occurred in daylight and 82% on the dry road surface. The mode choice for a cyclist is affected by the adverse environmental conditions. It is established in the literature [15] that cyclist's mode choice is highly varied and susceptible to change due to change in the environmental conditions. As the mode usage during these adverse conditions is low, therefore, reported crashes are also low in these conditions. There is a strong bias towards daylight crashes. This has the potential to result in modelling inaccuracy in the predictive deep learning model, as it will be difficult for the neural network to learn, classify and test effectively, and distinguish between different output variables. Therefore, lighting variables are further grouped into another environmental variable, i.e., meteorological road surface condition (Table VI).

Table V. Crash recorded in a) classification by fatality b) varying environmental conditions in the study area.

Time Period	2005-2018	Variable	Value
Slight	2638	Darkness	542(16.8%)
Serious	661	Daylight	2683 (83.2%)
Fatal	26	Dry	2644 (82%)
Total	3325	Wet/Frost/Snow	581(18%)

For understanding the spatial variation of crashes and hypothesis testing, following heatmaps are constructed.

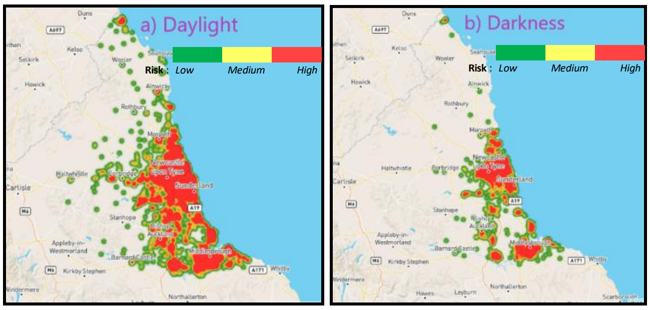


Fig. 2. Heat maps for hotspot identification for a) Daylight, and b) Darkness

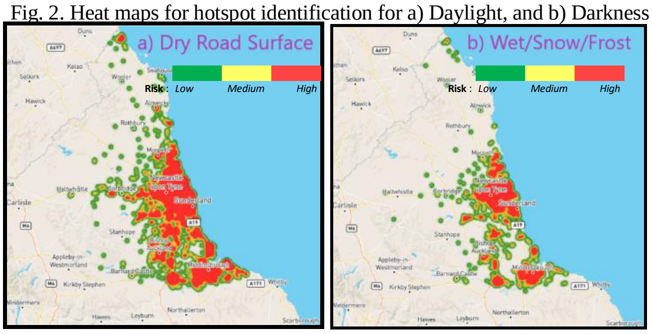


Fig. 3. Heat maps for hotspot identification for a) Dry Road Surface, and b) Wet/Snow/Frost.

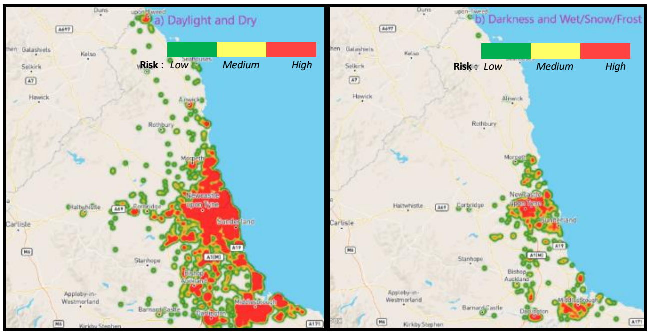


Fig. 4. Heat maps for hotspot identification for a) Daylight, and Dry Road Surface, and b) Darkness, and Wet/Snow/Frost.

Table VI. Crash recorded in the combined environment variable.

Variable	*Frequency (f)*	*%age*	*Variable*	*f*	*%age*
Darkness - No Street Lighting, and Dry	25	0.8	Darkness - Street Lights present, unlit and Dry	*8*	*0.2*
Darkness - No Street Lighting, and Wet/Damp	23	0.7	Darkness - Street Lights present, unlit and Wet/Damp	*1*	*0*
Darkness - Street Lighting Unknown, and Dry	10	0.3	Daylight and Dry	*2333*	*72.3*
Darkness - Street Lighting Unknown, and Wet/Damp	4	0.1	Daylight and Frost	*16*	*0.5*
Darkness - Street Lights present, lit and Dry	268	8.3	Daylight and Snow	*1*	*0*
Darkness - Street Lights present, lit and Snow	6	0.2	Daylight 0and Wet/Damp	*333*	*10.3*
Darkness - Street Lights present, lit and Wet/Damp	197	6.1	*n/a*	*n/a*	*n/a*

The hotspot results suggest that how different locations act as a hotspot depends on the environment variables of lighting and meteorological road surface condition. There is an expected centralization in Newcastle city centre, as it has a higher cyclist flow than the rest of the study area. Similar results for the city centre have been reported in the literature for university towns (see [55]). Newcastle, predominantly a university town, is surrounded by two big universities. It is evident from Fig 2-4 that except for this centralization; the crash pattern in terms of both frequencies as well as density and spread is significantly different for each environmental variable case. The study area is made up of local council units (Fig 1) with varying land-use patterns and traffic mixes in terms of composition and 85th percentile speed. As a result, the design and operation of infrastructure networks alter (modelled in the predictive infrastructure model), and their safety is influenced by changing environmental circumstances. This leads to conclude that the safe usage of infrastructure depends on the environmental conditions that a cyclist is subjected: i) Daylight and darkness, ii) Dry, and wet/snow/frost road surface, and iii), Daylight with dry, and darkness with wet/ snow/frost road surface condition. These conditions result in a varied level of risk for the same type of infrastructure to the cyclist, making the rider's subjected environmental conditions a dynamic road safety variable.

This is an unexpected finding, compared with the traditional road safety models/theories available in the literature. The present models although acknowledge, that road infrastructure and safety are interlinked. However, they do not consider that environmental conditions affect different infrastructure's safety in a varied manner, reinforcing the conclusion from Dublin cycling model [44], that the present models do not consider the cyclist's limitations and vulnerability. Unlike cyclists, the motorists are not adversely affected by these adverse environmental conditions, e.g., wet road surface condition will only affect the friction and skid resistance for the motorists. This effect is usually the same across all types of infrastructure. However, for a cyclist, the interaction with the infrastructure is already much more complicated and difficult. These adverse conditions pose varying challenges for the rider while using the infrastructure, which results in both physical and cognitive strains, and therefore act as a significant road safety variable (safety law of cognitive capability [12]). Thus, complex environmental conditions of lighting and meteorological road surface condition, alone and in combination with each other affect the cyclist's safe interaction with infrastructure. This variable needs to be modelled effectively and efficiently to develop the requisite knowledge-driven approach for cycling infrastructure. Such modelling will allow for the development of a dynamic safety index, which will allow safety analysis at the individual level (nanoscopic), rather than at a macroscopic level. To estimate the safety at an aggregate level, such as the city, this nano-safety can be aggregated to depict an area's overall safety. This shift in safety analysis towards nanoscopic modelling can help achieve the zero-vision road traffic fatality, demonstrated in the next section.

B. Deep Learning model.

The deep learning predictive models are constructed with a highly non-linear structure comprising of two hidden layers; with the principle characteristics described in Table VII. To evaluate the accuracy of the models, ROC curves are developed (Fig 5), and for numerical quantification, AUROC values are presented in Table VIII. In addition to the average AUROC values for three models, each output variable's values are also tabulated to ensure that both the overall accuracy and individual accuracy of each variable are evaluated. To compare the model's performance with the probability-based statistical model, lift charts are developed (Fig 6).

There are three different predictive models constructed using the input i) Spatial, ii) Personal, and iii) Infrastructure variables. Significantly high accuracy is obtained in all these constructed models, with the output of the 'riskiest environmental subgroup'. The model can take 13 output values; therefore, the ideal 100% accurate model will have an AUROC value of 13. The following AUROC values are obtained for, i) Spatial: 12.36 (95%), ii) Personal: 10.22 (79%), and iii) Infrastructure variables: 11.11 (85%), with an average value of 11.23 (86%). The accuracy achieved for the least accurate model (i.e. personal attribute) is significantly higher compared to available models in the literature (e.g. [45] found an error of more than 2/3 in Finnish TRAVA safety model for a cyclist, [56] found that due to inaccuracy, around 70% of the European countries either don't or rarely use crash prediction models). Although the present models in literature mostly serve their intended purpose, however for the cyclist these need to take into the specific safety variable such as the variable environmental conditions. The individual prediction capability of each of the 13 subgroups that the output can take is also evaluated separately. The median prediction accuracy of these 13 outputs for spatial, personal, and environmental model is 98%, 87% and 79% respectively. Thereby establishing the credibility of the constructed model. The high accuracy is attributed to the ability of deep learning methodology to model the non-linear complex relationships. The crashes are multifactored and relationship between the contributory factors is highly non-linear and complex. Therefore, it can be concluded based upon the accuracy obtained, that deep learning is a useful methodology for road safety investigation to develop accurate and efficient nanoscopic safety models.

Table VII. Characteristics and Structure of the constructed Network

Characteristics and Structure of the constructed Network				
Input Layer		***Spatial model***	***Personal Model***	***Infrastructure model***
	1	Hour	Gender	Road Type
	2	Number of Vehicles	Age	Speed limit
	3	Month	Age and Gender (combined)	1st Road Class
	4	Day	Journey Purpose	Road Hierarchy Level
	5	Weekday or Weekend	n/a	Road Hierarchy level and direction
	6	n/a	n/a	Junction Detail
	7.	n/a	n/a	Junction Control
	8	n/a	n/a	2nd Road Class
	9	n/a	n/a	Vehicle Maneuver
	10	n/a	n/a	Vehicle Junction Location
	11	n/a	n/a	Road Location of vehicle
	12	n/a	n/a	Carriageway Hazards
	No. of Input Units			50/ 29/86
Hidden Layer(s)	Total No. of Hidden Layers			2
	Total No. of Units in the Hidden Layers			700 (350in each layer)
Output Layer	Dependent Variables			Riskiest Environment Condition
	Total No. of Output units			13
	Error Function			Cross-entropy
Activation Function for Hidden Layers				Hyperbolic tangent
Activation Function for Output Layer				SoftMax

Table VIII. The area under the curve for the three constructed deep learning models

	Spatial	**Personal**	**Infra-structure**	**Average**
Darkness - No Street Lighting, and Dry	0.94	0.74	0.87	0.85
Darkness - No Street Lighting, and Wet/Damp	0.92	0.81	0.97	0.9
Darkness - Street Lighting Unknown, and Dry	0.7	0.88	0.96	0.85
Darkness - Street Lighting Unknown, and Wet/Damp	1	0.94	0.82	0.92
Darkness - Street Lights present, lit and Dry	0.98	0.67	0.86	0.84
Darkness - Street Lights present, lit and Snow	1	0.97	1	0.99
Darkness - Street Lights present, lit and Wet/Damp	0.99	0.75	0.87	0.87
Darkness - Street Lights present, unlit and Dry	0.98	0.66	0.87	0.84
Darkness - Street Lights present, unlit and Wet/Damp	1	0.85	0.42	0.76
Daylight and Dry	0.96	0.64	0.84	0.81
Daylight and Frost	0.96	0.91	0.87	0.91
Daylight and Snow	1	0.79	0.9	0.9
Daylight and Wet/Damp	0.92	0.63	0.86	0.8
Total	12.36	10.22	11.11	11.23
Average	0.95	0.79	0.85	0.86
Median	0.98	0.87	0.79	
high low				

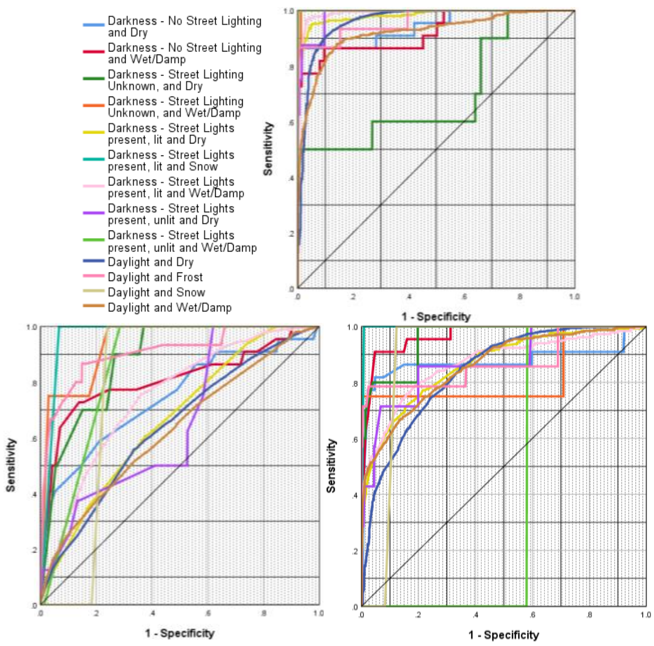
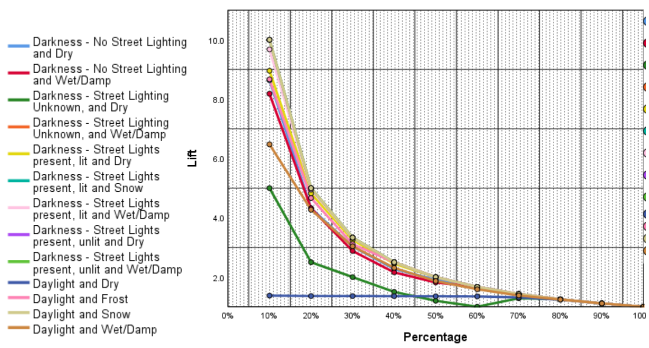


Fig. 5. Receiver Operating Characteristics (ROC) curve a) Spatial, b) Personal, and c) Infrastructure



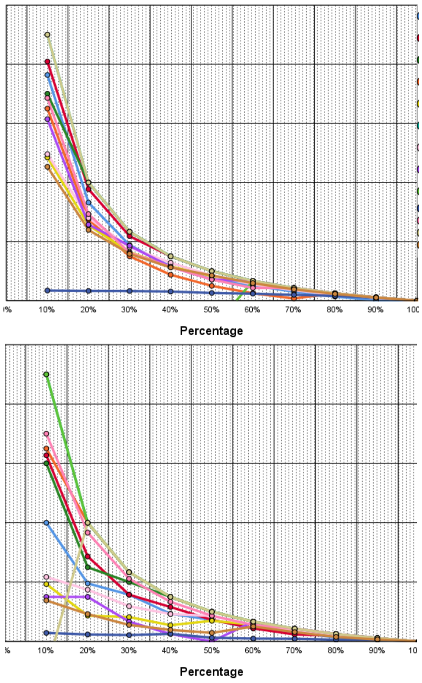


Fig. 6. Lift Chart a) Spatial, b) Personal, and c) Infrastructure

C. Causal Inference Method

1) Importance of input variables

The importance of each variable and their respective normalized importance are calculated and tabulated in Table IX.

Table IX. Normalized importance of various variables in the three constructed models

Variable		**Importance**	**Normalized Importance**
Spatial	Hour	0.281	100.0%
	Number of Vehicles	0.168	59.7%
	Month	0.264	93.7%
	Day	0.179	63.8%
	Weekday or Weekend	0.108	38.2%
Personal	Driver Gender	0.166	48.2%
	Driver Age Group	0.281	81.5%
	Age and Gender	0.345	100.0%
	Journey Purpose of Driver/Rider	0.208	60.3%
Infrastructure	Road Type	0.083	82.8%
	Speed Limit	0.085	84.8%
	1st Road Class	0.073	73.0%
	Road hierarchy level	0.076	76.0%
	Road hierarchy level and direction	0.092	91.2%
	Junction Detail	0.086	85.6%
	Junction Control	0.070	69.4%
	2nd Road Class	0.089	88.9%
	Vehicle Manoeuvre	0.101	100.0%
	Carriageway Hazards	0.078	77.1%
	Road Location of Vehicle	0.084	83.3%
	Junction Location of Vehicle	0.082	81.3%

In the spatial model, the most critical variables are hour and month in which the journey is undertaken. The hour and month of travel can also be correlated with the lighting conditions, however, more than 80% of the crashes have occurred in daylight. Both these variables are a representation of the traffic flow regime. Traffic flow is reported a significant variable affecting cycling safety (see [57]–[59]). The number of vehicles that are involved in the crash is not a significant variable, however, the overall traffic flow regime that the cyclist is exposed to during the entire trip is found to be a critical variable. The safety is negatively affected as the number of variables to be considered by the cyclist increases (safety law of cognitive capability [12]). This increase in the variable during the entire trip imparts the unsafeness in the interaction. This leads to the conclusion that traffic flow regime directly impacts the probable riskiest environmental condition to be experienced by the cyclist. A more in depth understanding of these variables in future has the potential to lead towards development of a real time autonomous route selection model, based upon the prevalent flow regime and environmental conditions.

In the personal variable model, the most important variable is the age and gender (combined) (100%), followed by driver age (81%), journey purpose (60%), and gender of the trip maker (48%). The rider's age and gender have an impact on how they react to varying environmental conditions (vulnerability, and experience of different age groups). The rider belonging to different age and gender have varied physical and cognitive abilities, thereby reacting differently to varied adverse environmental conditions. The result is a contribution to the understanding of how personal attributes of the rider affect their safety. Although age is an expected variable, however, results have shown that both gender and age in combination with each other (however, gender alone being the least significant variable), affect safe usage of the infrastructure in varied environmental conditions.

In Infrastructure model, the most critical variable is the vehicle manoeuvre (100%), followed by road hierarchy level and direction (91%), second road class (89%), junction detail (86%), and speed limit (85%). The least important variable is the control employed at the junction. This leads us to infer that the environmental conditions become critical when the cyclist must perform specific manoeuvres while interacting in the natural road environment. This is followed by the difference in the road hierarchy level and the corresponding direction of change in the hierarchy of road networks in which the cyclist is required to perform these specific manoeuvres. The third variable is the second road class. The road hierarchy level and direction, and the second road class are correlated with each other. The variable of road hierarchy level and direction signifies the difference between the first and second functional road classes. The next important variable is junction details and speed limit. Therefore, we can conclude, at intersections, environmental conditions become critical based upon the specific riding manoeuvres, the difference in road hierarchy level and direction of the change in road hierarch, and junction details. These are the most critical infrastructure parameters, affecting safe usage of the infrastructure under varying environmental conditions.

The novel variable introduced in this research, i.e., road hierarchy level and direction is found to be a critical variable. This can be attributed to a sudden change in driver behaviour, infrastructure parameters, and change in traffic flow regimes (which has been found critical in the spatial model). The motorists are not affected by such scenarios, as they are required by law to change the speed (with the change in the road hierarchy) and adhere to the speed limit on specific roads. The cyclist needs to make an immediate change in its riding style, the relative safety margin of errors, and its manner of interaction with the motorists. The motorist may start sudden accelerations, as they may want to accelerate suddenly if they have moved to a higher hierarchical functional road class, negatively affecting its interaction with the cyclists. The design elements of roadway also change drastically due to a change in the road hierarchy (see [60], [61]). The cyclist is more susceptible to these changes, whereas these infrastructure elements are designed specifically for the motorists, and their expected manoeuvres. Therefore, the research reinforces a requirement for planning and designing the infrastructure to move towards a more holistic approach while considering this vulnerable road user's limitations. Suppose we are to achieve a sustainable urban transport system. In that case, the cycling mode share has to increase by many folds ([62] highlighted the importance of increasing modal share of the cyclist in their scenario analysis for a sustainable transport system for the study area). This increase can only be achieved, if we make cyclist the pivot of our infrastructure design and network planning.

2) Statistical validation

The association between the target variable and input variables is tested statistically using Chi-square test. Their strength of the association with the riskiest environment is determined using Cramer's V value and Cohen's table.

Table X. Chi-square test for different variables across gender

Variable	**Dof**	**Chi-square value**	**p-value**	**H**	**Cramer's V**	**A**
Hour	13	2488.8	0.01	H_1	0.24	M
Number of Vehicles	4	130.1	0.01	H_1	0.1	S
Month	11	1080.9	0.01	H_1	0.18	M
Day	6	163.2	0.01	H_1	0.09	M
Weekday or Weekend	1	14.6	0.33	H_0	n/a	n/a
Driver Gender	1	24.1	0.03	H_1	0.09	S
Driver Age Group	6	267.0	0.01	H_1	0.11	M
Age and Gender	13	402.3	0.001	H_1	0.10	S
Journey Purpose of Driver/Rider	5	233.9	0.01	H_1	0.12	S
Road Type	5	80.3	0.01	H_1	0.07	S
Speed Limit	5	348.2	0.01	H_1	0.15	M
1st Road Class	4	167.6	0.01	H_1	0.11	S
Road hierarchy level	4	163.4	0.01	H_1	0.11	S
Road hierarchy level and direction	8	225.4	0.01	H_1	0.09	S
Junction Detail	8	342.4	0.01	H_1	0.12	S
Junction Control	3	164.3	0.01	H_1	0.13	S
2nd Road Class	5	241.6	0.01	H_1	0.12	S
Vehicle Manoeuvre	13	311.1	0.01	H_1	0.09	S
Carriageway Hazards	4	144.5	0.01	H_1	0.11	S
Road Location of Vehicle	7	127.1	0.07	H_0	n/a	n/a
Junction Location of Vehicle	8	206.8	0.01	H_1	0.09	S

where Dof is the degree of freedom, H is the hypothesis adopted, H_o; Null hypothesis: Interaction in the risky environment is independent of the variable, H_1; Alternate Hypothesis: Interaction in the risky environment is dependent on the variable, and A is the type of the association (S = Small, M = Medium, n/a = no association)

These results have depicted that the identified critical variables from deep learning model are associated with the risky environment at a 99.9% confidence interval. This is further validated by the Cramer's V value and the corresponding interpretation using Cohen's table. The only two variables, i.e., weekday or weekend, and road location of the vehicle, are not statistically associated; the same result from deep learning variable importance. Thereby validating the deep learning results statistically and developing the requisite confidence for model application and policy implications.

IV. Limitations

The study uses the crash database, based upon the reported crashes. However, there is a reported underreporting in the literature, especially concerning single cyclist slight crashes. In contrast, severe and fatal crashes are almost certainly reported due to the nature of the injury sustained. However, there are very few alternatives to using the crash database; although other methods have been explored, such as naturalistic study (see [63], [64]). However, these methods are still in infancy as their results cannot be quantified in terms of lives saved, or disruptions to the transportation network. Another methodology explored is making use of hospital data; however, such data cannot be further linked to the exact infrastructure location, time of the crash, and the prevalent traffic flow regime. The primary motivation for using the crash database is the ability to quantify the results and establish confidence for the policy implications, and further use of knowledge-driven measures by the road safety professionals.

V. Conclusion

In this work, the factors that determine the manner by which a cyclist interact with the road infrastructure is investigated, including variable environmental conditions of lighting and meteorological road surface. There are very few works in the literature which have modelled this variable. In the present literature, there have been compromises due to research focused on a single framework. The crashes are a multi-dimensional and multifactored phenomenon, requiring a similar multi-dimensional approach. In this study, an intelligent hybrid modelling framework is applied on north-east of England, consisting of a data collection unit and a knowledge processing unit (KPU). The KPU consists of: a) The statistical framework, b) Deep learning, and c) Causal inference. Three nanoscopic safety models have been constructed in the KPU, and a causal relationship has been identified between i) Spatial, ii) Infrastructure, and iii) Personal variables, with varying environmental conditions, followed by statistical validation. By combining multiple frameworks, we have demonstrated that a road safety model can be constructed with significantly high accuracy (spatial 95%, personal 79%, and infrastructure 85%, with an average of 86% across all models) and predictive power. To estimate the safety at the city level, this nano-safety can be aggregated to depict the overall safety of an area. A novel infrastructure variable, i.e., 'road hierarchy level and direction' is introduced in this study, which has been found critical. It is recommended that this variable is considered in cycling infrastructure planning and network design. The following main conclusions are deduced from the study a) Unsafeness of the interaction between user and infrastructure is dependent upon lighting and road surface meteorological conditions, b) Different environmental conditions pose different risks to different types of infrastructure, c) The riskiest environmental conditions are significantly affected by rider's gender and age group, and the prevalent traffic flow regime, d) The environment conditions significantly affect the interactions in which the rider needs to undertake specific manoeuvres due to a sudden change in the road hierarchy. The change in road hierarchy level and direction of change, (i.e., from higher hierarchical functional infrastructure type to a smaller one or vice versa) impacts the safety interactions, and e) The increase in the number of variables that are encountered during the entire trip negatively impact cycling safety.

A shift in the road safety analysis towards nanoscopic modelling can help achieve zero-vision road traffic fatality. The research reinforces a need for planning and design of infrastructure to move towards a more holistic approach while considering the limitations of this vulnerable road user. The result can contribute towards improving road safety and lead towards the development of a sustainable integrated cycling transportation system. It is hoped that this research will help reduce the cyclist crashes, thereby contributing to the promotion of this travel mode. The final output variable, i.e., the riskiest environmental condition, can be correlated with many underlying factors. Therefore, future research should aim to create a heterogeneous model, which can uncover the underlying variables.

Acknowledgement: The authors would like to thank the Gateshead City Council for providing access to the TADU crash database, and Edinburgh University for providing access to digimap.

REFERENCES

[1] J. Parkin, *Designing for Cycle Traffic : International principles and pactice*. London, United Kingdom: Institution of Civil Engineers, 2018.

[2] R. Aldred and S. Crosweller, "Investigating the rates and impacts of near misses and related incidents among UK cyclists," *J. Transp. Heal.*, vol. 2, no. 3, pp. 379–393, 2015.

[3] R. Elvik, "The non-linearity of risk and the promotion of environmentally sustainable transport," *Accid. Anal. Prev.*, vol. 41, no. 4, pp. 849–855, 2009.

[4] A. Laureshyn, M. De Goede, N. Saunier, and A. Fyhri, "Cross-comparison of three surrogate safety methods to diagnose cyclist safety problems at intersections in Norway," *Accid. Anal. Prev.*, vol. 105, pp. 11–20, 2017.

[5] N. Akgun, D. Dissanayake, N. Thorpe, and M. C. Bell, "Cyclist casualty severity at roundabouts - To what extent do the geometric characteristics of roundabouts play a part?," *J. Safety Res.*, 2018.

[6] M. Wardman, R. Hatfield, and M. Page, "The UK national cycling strategy: Can improved facilities meet the targets?," *Transp. Policy*, vol. 4, no. 2, pp. 123–133, 1997.

[7] J. Zhang, J. Lindsay, K. Clarke, G. Robbins, and Y. Mao, "Factors affecting the severity of motor vehicle traffic crashes involving elderly drivers in Ontario," *Accid. Anal. Prev.*, vol. 32, no. 1, pp. 117–125, 2000.

[8] A. Theofilatos, D. Graham, and G. Yannis, "Factors Affecting Accident Severity Inside and Outside Urban Areas in Greece," *Traffic Inj. Prev.*, vol. 13, no. 5, pp. 458–467, 2012.

[9] F. Bella and A. Calvi, "Effects of Simulated Day and Night Driving on the Speed Differential in Tangent-Curve Transition: A Pilot Study Using Driving Simulator," *Traffic Inj. Prev.*, vol. 14, no. 4, pp. 413–423, 2013.

[10] M. Sabir, J. van Ommeren, M. J. Koetse, and P. Rietveld., "Impact of Weather on Travel Demand and Mode Choice: An Empirical Analysis," in *Transatlantic NECTAR (Network on European Communications and Transport Activities Research) Conference*, 2009.

[11] J. Parkin, M. Wardman, and M. Page, "Estimation of the determinants of bicycle mode share for the journey to work using census data," *Transportation (Amst).*, vol. 35, no. 1, pp. 93–109, 2008.

[12] R. Elvik, "Laws of accident causation," *Accid. analyisis preention*, vol. 38, no. 4, pp. 742–747, 2006.

[13] D. Jaroszweski and T. McNamara, "The influence of rainfall on road accidents in urban areas: A weather radar approach," *Travel Behav. Soc.*, vol. 1, no. 1, pp. 15–21, 2014.

[14] J. Hong, D. Philip McArthur, and J. L. Stewart, "Can providing safe cycling infrastructure encourage people to cycle more when it rains? The use of crowdsourced cycling data (Strava)," *Transp. Res. Part A Policy Pract.*, vol. 133, no. January, pp. 109–121, 2020.

[15] E. Heinen, K. Maat, and B. van Wee, "Day-to-Day Choice to Commute or Not by Bicycle," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2230, pp. 9–18, 2011.

[16] F. Malin, I. Norros, and S. Innamaa, "Accident risk of road and weather conditions on different road types," *Accid. Anal. Prev.*, vol. 122, no. October 2018, pp. 181–188, 2019.

[17] A. Perrels, A. Votsis, V. Nurmi, and K. Pilli-Sihvola, "Weather conditions, weather information and car crashes," *ISPRS Int. J. Geo-Information*, vol. 4, no. 4, pp. 2681–2703, 2015.

[18] A. Alhasan, I. Nlenanya, O. Smadi, and C. A. MacKenzie, "Impact of pavement surface condition on roadway departure crash risk in Iowa," *Infrastructures*, vol. 3, no. 2, 2018.

[19] A. R. Lawson, V. Pakrashi, B. Ghosh, and W. Y. Szeto, "Perception of safety of cyclists in Dublin City," *Accid. Anal. Prev.*, vol. 50, pp. 499–511, 2013.

[20] C. Hydén, "The development of a method for traffic safety evaluation: the Swedish traffic conflict technique. Doctoral thesis.," Lund University, 1987.

[21] A. Laureshyn and A. Varhelyi, "The Swedish Traffic Conflict Technique: observer's manual.," 2018.

[22] R. Aldred and A. Goodman, "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.*, vol. 110, no. June 2017, pp. 161–170, 2018.

[23] K. Schleinitz, T. Petzoldt, L. Franke-Bartholdt, J. Krems, and T. Gehlert, "The German Naturalistic Cycling Study – Comparing cycling speed of riders of different e-bikes and conventional bicycles," *Saf. Sci.*, vol. 92, pp. 290–297, 2017.

[24] B. Schreck, "Cycling and designing for cyclists in Germany: Road safety, Guidelines and Research," *Trans. Transp. Sci.*, vol. 8, no. 1, pp. 44–57, 2017.

[25] TRL, "Infrastructure and Cyclist Safety: Research Findings TRL Report PPR 580," 2011.

[26] M. E. Campbell, "The wet-pavement accident problem: Breaking through.," *Traffic Q.*, vol. 25, no. 2, pp. 209–214, 1971.

[27] K. Smith, "How seasonal weather conditions influence road accidents in Glasgow.," *Scott. Geogr. Mag.*, vol. 98, no. 2, pp. 103–114, 1982.

[28] S. . Changnon, "Effects of summer precipitation on urban transportation," *Clim. Change*, vol. 32, pp. 481–494, 1996.

[29] J. Andrey and R. Olley, "Relationships between weather and road safety, past and future directions," *Climatol. Bull.*, vol. 24, no. 3, pp. 123–137, 1990.

[30] J. Andrey, B. Mills, M. Leahy, and J. Suggett, "Weather as a chronic hazard for road transportation in Canadian cities," *Nat. Hazards*, vol. 28, no. 2–3, pp. 319–343, 2003.

[31] K. Keay and I. Simmonds, "The association of rainfall and other weather variables with road traffic volume in Melbourne, Australia," *Accid. Anal. Prev.*, vol. 37, no. 1, pp. 109–124, 2005.

[32] L. Qiu and W. Nixon, "Effects of Adverse Weather on Traffic Crashes: Systematic Review and Meta-Analysis," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2055, no. 1, pp. 139–146, 2008.

[33] D. Hambly, J. Andrey, B. Mills, and C. Fletcher, "Projected implications of climate change for road safety in Greater Vancouver, Canada," *Clim. Change*, vol. 116, no. 3–4, pp. 613–629, 2013.

[34] M. Jackett and W. Frith, "Quantifying the impact of road lighting on road safety - A New Zealand Study," *IATSS Res.*, vol. 36, no. 2, pp. 139–145, 2013.

[35] C. F. Lio, H. H. Cheong, C. H. Un, I. L. Lo, and S. Y. Tsai, "The association between meteorological variables and road traffic injuries: A study from Macao," *PeerJ*, vol. 2019, no. 2, pp. 1–13, 2019.

[36] M. S. Obeidat, S. K. Khrais, B. S. Bataineh, & Majd, M. Rababa, and M. M. Rababa, "Impacts of roadway lighting on traffic crashes and safety in Jordan," *Int. J. Crashworthiness*, 2020.

[37] O. A. Kasm, Z. Ma, J. Y. J. Chow, and A. Diabat, "Quantifying the effect of cyclist behavior on bicycle crashes and fatalities," in *98th Annual Meeting of the Transportation Research Board*, 2019.
[38] G. Lovegrove and T. Sayed, "Using Macro-level Collision Prediction Models in Road Safety Planning Applications," *Transp. Res. Rec. J. Transp. Res. Board,* vol. 1950, pp. 73–82, 2006.
[39] J. Ambros, C. Jurewicz, S. Turner, and M. Kiec, "An international review of challenges and opportuniy in development and use of crash prediction models," *Eur. Transp. Researh Rev.*, 2018.
[40] D. Lord and F. Mannering, "The statistical analysis of crash-frequency data : A review and assessment of methodological alternatives," *Transp. Res. Part A,* vol. 44, no. 5, pp. 291–305, 2010.
[41] F. L. Mannering and C. R. Bhat, "Analytic Methods in Accident Research Analytic methods in accident research : Methodological frontier and future directions," *Anal. Methods Accid. Res.*, vol. 1, pp. 1–22, 2014.
[42] D. Gettman, L. Pu, T. Sayed, and S. Shelgy, "Surrogate Safety Assessment Model and Validation : Final Report, Report No. FHWA--HRT-08-051," 2008.
[43] J. C. Calvey, J. P. Shackleton, M. D. Taylor, and R. Llewellyn, "Engineering condition assessment of cycling infrastructure: Cyclists' perceptions of satisfaction and comfort," *Transp. Res. Part A Policy Pract.*, vol. 78, pp. 134–143, 2015.
[44] A. Lawson, "An Analysis of Cycling Safety and development of a bicycle trip assignment methodology," Trinity College Dublin, 2015.
[45] H. Peltola and R. Kulmala, "Accident models," Espoo, Finalnad, 2010.
[46] D. for T. DfT, "Total Road length (miles) by road type and local authority in Great Britain, RDL01012a.," London, United Kingdom, 2019.
[47] F. Mannering, C. R. Bhat, V. Shankar, and M. Abdel-aty, "Analytic Methods in Accident Research Big data , traditional data and the tradeoffs between prediction and causality in highway-safety analysis," *Anal. Methods Accid. Res.*, vol. 25, p. 100113, 2020.
[48] IBM, "IBM SPSS Neural Networks 25," New York, USA, 2017.
[49] K. Gharehbaghi, "Artificial neural network for transportation infrastructure systems," *MATEC Web Conf.*, vol. 81, 2016.
[50] S. Haykin, *Neural Networks, A Comprehensive Foundation*, Second Edi. Pearson Education (Singapore) Pte.Ltd, 2005.
[51] M. G. Karlaftis and E. I. Vlahogianni, "Statistical methods versus neural networks in transportation research: Differences, similarities and some insights," *Transp. Res. Part C Emerg. Technol.*, vol. 19, no. 3, pp. 387–399, 2011.
[52] Simon Haykin, *Neural Networks and Learning Machines*, 3rd editio. Prentice Hall, Inc., 2014.
[53] Environment and Transport Overview and Scrutiny Committee, "Road Casualty Reduction in Leicestershire 2000 to 2014," London, United Kingdom, 2015.
[54] J. . Cohen, *Statistical power analaysis for the behavioral science*, 2nd ed. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1998.
[55] B. Gatersleben and K. M. Appleton, "Contemplating cycling to work; attitudes and perceptions in different stages of change," *Transp. Res. Part A*, vol. 41, pp. 302–312, 2007.
[56] G. Yannis, A. Dragomanovits, A. Laiou, T. Richter, S. Ruhl, and F. Calabretta, "Inventory and critical review of the existing APM's and CMF's and related data sources," Confederation of European Directors of Roads, Brussels, Belgium, 2015.
[57] K. Stewart and A. McHale, "Cycle lanes: their effect on driver passing distances in urban areas," *Transport*, vol. 29, no. 3, pp. 307–316, 2014.
[58] J. Parkin, M. Wardman, and M. Page, "Models of perceived cycling risk and route acceptability," *Accid. Anal. Prev.*, vol. 39, no. 2, pp. 364–371, 2007.
[59] TfL, "London Cycling Design Stnadards," London, United Kingdom, 2016.
[60] DMRB TD9/93, "Highway Link Design : Road Geometry Links Volume 6 Section 1," 1993.
[61] . Highways England, "Cycle Traffic and the Strategic Road Network. Interim Advice Note 195/16," Guildford, UK, 2016.
[62] M. C. Bell CBE, F. Galatioto, C. Fryer, and R. Wang, "The role of cycling in delievering sustainable travel by 2050 .," in *48th Universities Transport Studies Group Annual Conference*, 2016.
[63] M. Dozza and A. Fernandez, "Understanding Bicycle Dynamics and Cyclist Behavior From Naturalistic Field Data," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 1, pp. 376–384, 2014.
[64] I. Walker, "Drivers overtaking bicyclists: Objective data on the effects of riding position, helmet use, vehicle type and apparent gender," *Accid. Anal. Prev.*, vol. 39, no. 2, pp. 417–425, 2007.