Intelligent nanoscopic road safety model for cycling infrastructure

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.ac.uk

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

***Abstract*— This paper is concerned with the development of intelligent safety modelling for cycling safety at the nanoscopic level. The present models are primarily focused on the motorists modelling at an aggregate level. In this work a framework for safety analysis is proposed consisting of a) Data collection unit, b) Data storage unit, and c) Knowledge processing unit. The predictive safety model is developed in the knowledge processing unit using supervised deep learning with neural network classifier, and gradient descent backpropagation error function. This framework is applied to a case study in Tyne and Wear county in England's northeast by using the crash database. An accurate safety model (88% accuracy) is developed with the output of the riskiest age and gender group, based upon the specific input variables. The most critical variables affecting the safety of an individual belonging to a particular age and gender groups, are the journey purpose, traffic flow regime and variable environmental conditions it is subjected to. It is hoped that the proposed framework can help in better understanding of cycling safety, aid the transportation professional for the design and planning of intelligent road infrastructure network for the cyclists.**

Keywords—intelligent transportation system, road safety models, infrastructure, deep learning

I. INTRODUCTION

There were 1,870 fatalities, 25,950 serious and 129,810 slight injuries due to road traffic crashes in Great Britain in 2019 [1]. Nationally, the road traffic collisions cost the UK economy more than £35 billion every year [2]. While cyclists account for only 2% of the trip share and only 1% of the distance travelled in Great Britain, they, however, face a disproportionate share of risk and causalities. In effect, the risk currently faced by cyclists is highest amongst any road user in Great Britain, 12.5 times higher than the motorist for the same traversed distance. In the European region, the percentage of cyclist fatalities has increased from 6% in 2007 to 8% in 2016. Therefore, the problem of improving the safety of cyclists to reduce the numbers of cyclist's fatalities is a primordial one requiring special attention.

The preference and requirements of cyclists are different from other road users [3]. Safety is the main barrier associated with this mode, which is a critical mode and route choice variable [4]. The cycling time spent in varying infrastructural and environmental conditions is an important variable [5] and is influenced by sociodemographic and work characteristics of the trip maker [6]. There are limited studies which explore the risk of cyclists to their exposure [7], and there is insufficient evidence to understand the relationship between cyclist safety, and the identified safety parameters [8]. Additionally, there is a need for the capabilities to assess the safety of the experimental roadway designs and operational strategies before they are built or employed in the field [9]. This can be achieved by constructing a dynamic safety model, which is based upon these identified variables rather than the present probabilistic function of the traffic flow.

An increase in the safety for cyclist will result in an increase in the cycling mode share. The real and perceived risks are the major barrier for the uptake of this sustainable mode of travel. The personal attributes of the rider have been reported as a significant variable which affects the safe usage of the infrastructure. In Czech Republic [10] (Bíl, Bílová and Müller, 2010), found that males account for around 69% of the crashes, and are more likely to be involved in a fatal crash (80%) [10]. Similarly (Rodgers, 1995), found that males are at a higher risk than females (around 5 times more likely than females for the same distance traversed) [11]. Similarly, the age of the rider significantly affects the safety of this mode, which is dominated by younger adults [12]. The study in England for the assessment of road safety by travel mode led them to conclude that risks for road users are highest in their youth. Their risks fall with the age. The similar results were also obtained in the Netherlands [13].

The study to understand the cyclist's injury by age and gender in Sweden concluded that the females show a lower incidence than males, however, the elder women are more likely to be involved in a serious crash, than the younger women. The same results have been reported for males, with even more difference between the young and elderly population. They found that females sustain more work trip injuries than men[14]. However, men are more reluctant for modal shift to cycling than women [15], and it takes much more improvement in the infrastructure and environment for the women to consider cycling [12].

At the present road, safety analysis is performed using fatality and injury rate. The sole usage of statistics is insufficient to achieve a thorough understanding of road safety and developments over time [17], [18]. The current modelling is based upon the complex human factors [19] [20], believed to be directly or indirectly responsible for most of the crashes. The output of these prediction models gives prediction over a long-term with the main aim to forecast the yearly crash, their seasonal variation and identification of the major black spots. These are primarily based upon the assumption that instantaneous traffic flow is the direct representation of the human factors responsible for the crashes. As the flow increase, the probability of the interaction increases and so does the probability of a crash [21]. All the major crash prediction models British [22], USA /Canada Model [23], Danish Model [24], Swedish Model [25], Finnish Model [26], etc. are all based on this assumption. However, from the literature, the cyclist is found to be susceptible to several parameters which other road users are not subjected to, such as the personal attribute of age and gender. Therefore, the present safety models are inable to model the cyclist safety effectively and efficiently.

Cycling can result in both cognitive and physical strains, to which riders of different age and gender will respond differently due to different physical and physiology capabilities. The motorists are also influenced by the appearance of the cyclist [6], which further complicates the interaction, thereby making it a major road safety variable. The modelling at present is focussed on the overall usage of the infrastructure primarily at an aggregate, rather than at microscopic level. The attempt to model such a variable is now possible due to the advancement in data-driven science. If we are to increase the cycling mode share, mitigation measures/ planning of infrastructure needs to be user-focussed at the microscopic level and then aggregation can take place to obtain the results at the city or a . The motorists have the advantage of being in a closed relatively safe environment. Also, the cities are randomly changing, the age and gender distribution of one city can be significantly different from the other, e.g. in the UK; a university town such as Oxford or Cambridge has significant different age distribution than the old English mining towns such as Sunderland. Similarly, the cities are now growing differently due to changing land-use-pattern, immigration, education institutes, etc. Therefore, with the changing patterns of the cities, planning needs to incorporate these through the development of Intelligent Transportation Systems, adhering to the specific needs of the city. This will result in more sustainable and smart cities, which will significantly improve urban liveability.

It is evident that presently there is a discrepancy between what is reported in the literature and what is practised by the professional. Therefore, the work presented here aims to develop a framework for developing a real-time Intelligent road safety model which can have direct implications for infrastructure planners/modellers. To achieve this, the following objectives are designed:

1. To develop a framework for road safety analysis.
2. To check the hypotheses that the safe usage of the infrastructure is dependent upon the personal attributes of the rider
3. To develop a road safety model.
4. Develop an understanding of the relationship between the identified input variables and safety
5. Identify the most important variables affecting safety.

In the next section, the proposed framework and the study area are defined (section II), followed by the methodology section (III). The results are presented in section IV, and the conclusions drawn in section V.

II. PRPOSED FRAMEWORK AND STUDY AREA

The following Intelligent modelling framework is proposed (Fig 1)

A. Data Collection Unit

The data collection unit will continuously collect the data concerning the safety of the modelled user. This can be from a variety of sources depending upon the type of investigation e.g. instrumented vehicle for a naturalistic study, the dash-board camera of the response vehicle, crash database, etc.

B. Data Transmission/ Storage Unit

The data collected will be transmitted to the main database/ server, either immediately through the internet or stored in a memory device and later transmitted to the server to ensure safe and secure storage. In this unit, the aggregation of data will be performed for further evaluation and modelling. As the data inflow is continuous, therefore new data transmitted will be used to constantly update the model, ensuring the final in-use model does not age with time.

C. Knowledge Processing Unit (KPU).

This raw data will be transformed into knowledge by the Knowledge Processing Unit (KPU) Fig 1. This will be undertaken by identifying the correlation between the input parameters, and then developing causation matrices. It will be performed by KPU through deep learning with neural network classifier, and gradient descent backpropagation error function. A predictive model will be constructed based upon the pre-defined attributes, which when given the input can predict the safety in real-time for the end-user. For each new dataset, the model will first train itself and then test to ensure the desired accuracy is achieved, followed by validation.

In this study, the framework is applied for the safe modelling of cycling infrastructure based upon the historic crash dataset (2005-2018) on the study area of Tyne and Wear county in the northeast of England (Fig. 2). The output modelled is the riskiest age and gender of the rider based upon the attributes of a) Spatial, b) Environmental, and c) Infrastructure variables. It is one of the nine official regions of England, encompassing an area of 3,317 sq. miles, housing five boroughs with a population of 1.13 million, and an estimated 693,000 jobs.

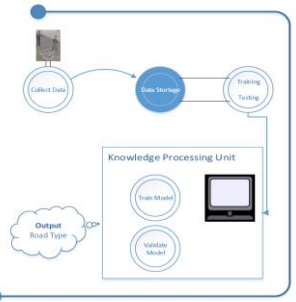

Figure 1: Knowledge Processing Unit (KPU).

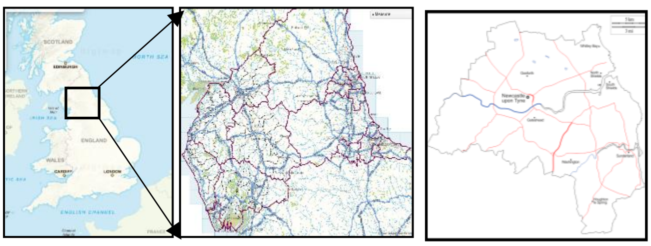

Figure 2: Location and Boundaries of the study area

III. METHODOLOGY

The detailed crash investigation for each crash is performed by the concerned local authorities. A trained road crash investigator visits crash site and records the requisite crash details in a pre-defined document, set out by the department for Transport known as "STATS 19", consisting of four sections, i) Accident Statistics, ii) Vehicle Record, iii) Casualty Record and iv) Contributory Factors. The attributes of each crash are recorded, i.e., i) type of severity, ii) Time, date and location of the crash, iii) Environment conditions such as lighting conditions, 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. These details are stored on an online platform, housed by the Department for Transport (DfT). For this study, we were provided access to the crash database Traffic and Data Unit (TADU) available with the Gateshead city council.

In the Knowledge Processing Unit, correlation, and causation is investigated, and a predictive model is developed by using deep learning with neural network classifier, and gradient descent backpropagation error function. It is the sub-group of a machine learning techniques based upon computational methodologies which imitate working of the human brain. The neural networks were introduced firstly in transportation research in the 1990s [27]. The infrastructure problems are characterized by interconnectivity between physical and tangible assets, required for developing and supporting the nation. The neural network has been widely applied as a data analytic method in transportation [28]. They are very generic, accurate, and convenient mathematical models, simulating the numerical model components [29]. This is due to their ability to work with the huge amount of the multi-dimensional data, modelling flexibility, learning, generalization ability, adaptability and good predictive ability [29]. The main motivation for using deep learning for modelling safety is that crashes are highly non-linear, and the modeller has no guidance from either theory or even dimensional analysis for modelling. Although there exist other algorithms and deep learning is not a new concept, however, its ability to solve the complex and the interchangeable system problems, which the transportation system is characterized by, is the main motivation for employing it [30].

EXPAND THIS ONE.

A learning algorithm is developed to divide the data set randomly into training (65%), testing (30%), and validation (5%). This division ensures enough dataset for learning, assessment of the trained model and relevance to untrained scenarios [28]. The following network structure is used to construct the model.

Table 1: Network structure of the deep learning model

Network Topology	Number of hidden layers	2
	Elements in each layer	30
	Activation function between the hidden layers	Hyperbolic Tangent
	Activation function between hidden and output layer	SoftMax
	Error function	Cross-entropy
Training	Type	Batch
	Error function	Scaled conjugate gradient
	Initial Lambda	0.0000001
	Initial Sigma	0.000001
	Initial Centre	0
	Initial offset	±0.001
Stopping and Memory Criterion	Steps (max) without a change in the error	999
	Training (max) time	999
	Training (max) epochs	999
	Relative change in the training error (min)	0.0001
	Relative change in the training error ratio (min)	0.001
	Cases to store in the memory (max)	999
Hidden layers	Total No. of Hidden Layers	2
	Elements in each layer	30

A four-step iterative backpropagation algorithm is used.

Step 1: Random weights are assigned to each weighted connection between the input and hidden, first and second hidden, and between the hidden and output layers).

For signal propagation within hidden layers, Hyperbolic tangent' activation function is used given by:

$$O_a = \tanh(S_a) = \frac{e^{S_a} - e^{-S_a}}{e^{S_a} + e^{-S_a}} \qquad (1)$$

O_a is the activation of the ath output neuron

The 'SoftMax' activation function is used between the hidden and output layer, given by:

$$O_a = \sigma(S_a) = \frac{e^{S_a}}{\sum_{k=1}^{m} e^{S_k}} \qquad (2)$$

m *is the number of output neurons*

These functions take real numbers as arguments and return real values [-1, +1].

Step 2: The error between the predicted output and target output is calculated through cross-entropy error function.

$$E = -\sum_{a=1}^{m} t_a \ln O_a \qquad (3)$$

O_a *is the actual output value of the output node* a,

t_a is the largest value a, and m is the number of output nodes

Step 3: The initial random synaptic weights are updated based upon the error obtained in step 2. In each epoch, the backpropagation algorithm calculates the gradient of the training error as

a) nodes between the input and hidden layer

$$\frac{\partial E}{\partial w_{ha}} = \sum_{a=1}^{m}(O_a - t_a)\, x_h w_{ha}(1 - x_h)x_b \qquad (4)$$

b) nodes between the output and hidden and layer

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

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

$$\Delta w_{bh} = -\gamma \frac{\partial E}{\partial w_{ha}} \qquad (6)$$

$$\Delta w_{bh+1} = w_{bh} + \Delta w_{bh} \qquad (7)$$

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

Step 4: Iteration (scaled conjugate gradient): The updating of weights is iterated until either the minimum change in the training error or the maximum number of these iterations (epochs) is achieved.

To evaluate the performance of the constructed models, Area Under the Curve (AUC) of the Receiver Operating

Characteristics (ROC) curve is used, considered an effective measure of the accuracy [31]. It is a plot between the true positive rate (sensitivity) and false-positive rate (1-specificity), which evaluates distinguishable power of the constructed model between the true safety, and riskiest age and gender group. The numerical value of the area under this curve (AUC) is a measure of the separability, i.e. it measures whether a risky scenario based upon the input values is correctly predicted risky or not.

After establishing the credibility and predictive power of the constructed model, the research also aims to develop an understanding of the relationship between the input variables and safety. Therefore, the importance of each of the variable in the prediction model is determined by evaluating the sensitivity of the model to the change in the input values. Besides, normalized importance concerning the most critical variable is also determined.

EXPAND HERE

IV. Results

The following outputs are obtained from the KPU.

A. Statistical

There are 3,325 bicyclist crashes recorded in the study area, 79.3 % slight, 19.9% serious and 0.8% fatal crashes. The age and gender distribution of the crashes are presented in Tables 2 and 3.

Table 2: Crash distribution across age groups.

Age	Frequency	Per cent	Cumulative Percent
17-24	537	16.2	16.2
25-34	494	14.9	31.0
35-44	347	10.4	41.4
45-54	251	7.5	49.0
55-64	115	3.5	52.5
Over 64	65	2.0	54.4
Under 17	1420	42.7	97.1
Unknown	96	2.9	100.0
Total	3325	100.0	

Table 3: Crash distribution across gender

Gender		Collision Severity			
		Fatal	Serious	Slight	Total
Female	Number	1	90	278	369
	Percentage	0.03%	2.71%	8.36%	11.10%
Male	Number	25	571	2360	2956
	Percentage	0.75%	17.17%	70.98%	88.90%

It is evident that the risk that cyclist's faces vary with their gender and age. Therefore, the normalized risk for each age groups is presented in Table 4. which also considers the miles traversed by each group.

EXPAND A BIT here

Table 4: Normalized risk across age and gender

	Male/Female		
Age group	Risk rate	Age group	Risk rate
0-16	1.46	40-49	1.02
17-20	0.27	50-59	0.81
21-29	1.67	60-69	0.77
30-39	1.11	70+	0.10

Therefore, based upon the statistical output, we can conclude that the risk for the cyclist is highly varied based upon its attribute of age and gender. The risk is highest for the cyclists in the early stage of their life, and male cyclist faces a disproportionately higher risk in their youth compared to females. The risk for females increases as they grow older.

B. Predictive model

To develop the predictive model, the following input variable regarding each crash is inputted in the developed deep learning framework.

Table 5: Input variable table for the predictive model

No.	Input Variable	Values
1.	Spatial	
a)	Month of Journey	Jan-Dec.
b)	Journey Day	Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday.
c)	Journey Weekday/ Weekend	Weekday. Weekend.
d)	Journey Hour	0-23.
e)	Number of vehicles	1-5.
f)	Journey Purpose	Commuting, work trip, School Journey by Pupil, taking pupil to school, other, Unknown.
2.	Environmental	
a).	Lighting conditions	Daylight /Darkness- No Street Lighting, Street Lighting Unknown, Street Lights present and lit, Street Lights present but unlit,
b).	Weather (Meteorological) conditions	Fine/Rain/Snow-with high winds, without high winds, fog, or Mist Hazard, Other.
c).	Road Surface Condition	Dry, Frost/ice, Wet/damp, Snow
3.	Infrastructure	
a)	Road Type	Dual Carriageway, One-way street, Roundabout, single carriageway, slip road,
b)	Speed limit	20-70
c)	1st Road Class	A, B, C, E, U
d)	Road Hierarchy Level	0-4
e)	Road Hierarchy level and direction	-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)	2nd Road Class	A, B, C, E, U
i)	Vehicle Maneuver	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 meters 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
	Output Variable	Risk gender and Age Group

The predictive model constructed, can accurately predict the riskiest age and gender group based upon the specific input variables. The KPU not only develops the models but also provides the evaluating matrix for evaluation by the end-user. This helps to develop confidence in the constructed model for its application by road safety professionals. The evaluation is performed through a) ROC Curve (Fig. 3), b) AUC value (Table. 7), c) Gain charts (Fig. 4), and d) Lift charts (Fig. 5).

The ROC curve is close to the top left-hand corner (optimum ideal hypothetical scenario), depicting a good overall prediction capability of the constructed model. The lift chart evaluates the benefit of using the model rather than a general probability model. A significant gain as evident from the gain chart is achieved by the model, which is also depicted in the cumulative gain achieved shown in the lift chart, in which the gain at 10% data points varies between 2-8, with an average gain of 6.5.

EXPAND A BIT MORE

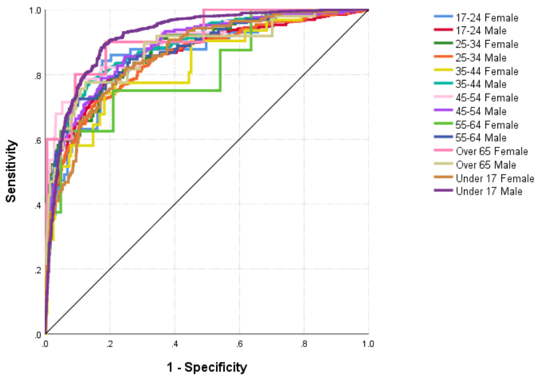

Figure 3:Receiver operating characteristic curve for the constructed model.

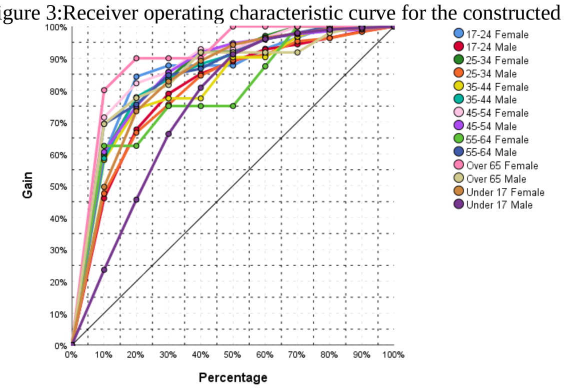

Figure 4: Gain chart for the constructed model

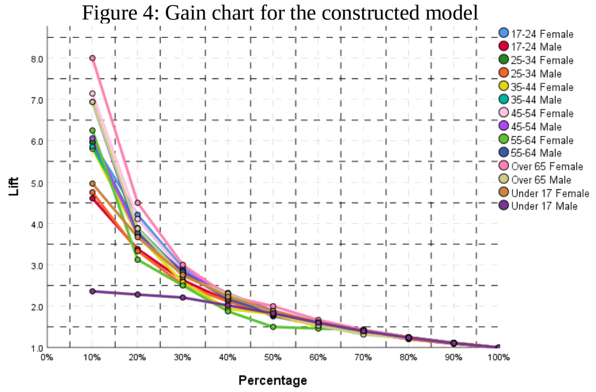

Figure 5: Lift curve for the constructed model

The accuracy of the models is evaluated through the prediction capability of each of the output variable, rather than the aggregate model only. This ensures that the predictive capability of one variable is not overrepresented and verifies that the model can estimate all the subgroups efficiently and accurately. The AUC values of the ROC curve for each output variable is presented in Table 6.

Table 6: Area under the ROC curve for different output variables

Area Under the ROC Curve			
Variable	Area	Variable	Area
Under 17 Female	0.86	35-44 Male	0.89
Under 17 Male	0.92	45-54 Female	0.90
17-24 Female	0.86	45-54 Male	0.88
17-24 Male	0.86	55-64 Female	0.82
25-34 Female	0.88	55-64 Male	0.88
25-34 Male	0.85	Over 65 Female	0.91
35-44 Female	0.83	Over 65 Male	0.88
Average	0.87	Standard Deviation	0.03
Median	0.88	Mode	0.86

Significantly high accuracy is obtained in the constructed model with average AUC value of 0.88 and a standard deviation of 0.03. This implies that the model can distinguish between the risky and non-risky scenarios in 88% of the presented scenarios. As the standard deviation is very low (3% of mean), we can therefore conclude the accuracy achieved is high across the output spectrum. Through inverse analysis, this leads us to infer that the risk which each sub-group faces is dependent upon the specific combination of input variables. This combination is specific to each subgroup of age and gender. The importance and normalized importance with respect to the most critical variable (journey purpose), for each of the input variable, is presented in Table. 7.

Table 7: Importance and Normalized importance values of the input variables

1.	Spatial	Importance	Normalized Importance
a)	Month of Journey	0.05	45%
b)	Journey Day	0.05	42%
c)	Journey Weekday/ Weekend	0.02	20%
d)	Journey Hour	0.06	55%
e)	Number of vehicles	0.05	41%
f)	Journey Purpose	0.11	100%
2.	**Environmental**		
a).	Lighting conditions	0.05	48.2%
b).	Lighting and road surface condition	0.06	50.9%
3	**Infrastructure**		
a)	Road Type	0.04	36.6%
b)	Speed limit	0.05	41.1%
c)	1st Road Class	0.04	35.5%
d)	Road Hierarchy level and direction	0.05	43.7%
e)	Junction Detail	0.05	42.1%
f)	Junction Control	0.04	31.9%
g)	2nd Road Class	0.04	37.4%
h)	Vehicle Maneuver	0.05	46.3%
i)	Road Location of vehicle	0.05	45.6%
j)	Vehicle Junction Location	0.05	45.4%
k)	Skidding and Overturning	0.03	29.1%

The most critical variables affecting the safety of the rider is the journey purpose, followed by the hour of the journey (a heterogeneous variable representing the traffic flow regime), environmental condition of lighting and road surface condition, and vehicle manoeuvres. These variables belong to different sub-groups, which reinforces the traditional road safety theory that the crashes are a multi-factor element

V. Summary and Conclusion

In this work, we have proposed a three-phase framework for road safety analysis of a) Data collection unit, b) Data transmissions/storage unit, and c) Knowledge processing unit (kpu). This framework is applied to a case study of the crash database in Tyne and Wear county in the northeast of England. Through the kpu, an accurate and efficient road safety model is developed with a high prediction capability to predict the most risk subgroup of age and gender, based upon the combination of input variables. We have proven the hypothesis that the safe usage of the infrastructure is dependent upon the personal attribute of the rider. It has been demonstrated that it is possible to predict the characteristic safety of the infrastructure for an individual using a) Spatial, b) Environmental, and c) Infrastructural parameters. The combination of these variables presents a specific risk to a specific population group.

The proposed model is validated using the available test data and an overall high level of accuracy (88%) is achieved. This small inaccuracy can be attributed to the dynamic nature of crashes. The results from the ROC curve, gain and lift charts suggest that the model can be employed for safety analysis with certainty. This proves the effectiveness of the proposed framework and develops the requisite confidence in the developed model for use by road safety professionals.

An understanding of the relationship between the identified input variables and safety have been developed. The most important variables affecting the safety of an individual is dependent upon the following variables in the descending order, a) The purpose of the journey, b) Traffic flow regime that is plying, and c) Prevalent environmental conditions. This reinforces that safety is a multi-factor element, which requires a dynamic approach. Therefore, based upon these input parameters we can assess the safety of the infrastructure. The results of the study can have a significant impact on the route choice, modelling and planning of infrastructure. Through the inverse analysis, the constructed model can assess with certainty regarding the type of infrastructure required to increase safety, thereby paving way for a knowledge-driven approach to cycling infrastructure.

HOW CAN THE MODEL RELATE WITH VEHICLE AUTOMATION.

Presently, we are at the doorstep of the fourth industrial revolution (autonomous transportation system), in which the route will be selected automatically by the autonomous system, therefore it is essential that the planning of transportation system also evolves, and real-time models are developed for city planners. These should be able to develop different measures/ optimize the infrastructure based upon its intended users and develop recommendation measures to increase safety (modal share) for a particular targeted population. The future direction of research should aim to develop a dynamic real-time road safety model.

We would like to thank Northumbria University for funding the research and Gateshead city council for providing access to crash database TADU.

References

[1] Department for Transport, "Road accidents and safety statistics," Her Majesty's Government, London, 2019.

[2] Department for Transport, *National Travel Statistics, Her Majesty's Government. About National Travel Survey*. 2019.

[3] 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, doi: 10.1016/j.aap.2016.04.035.

[4] 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, doi: 10.1016/j.jsr.2018.09.004.

[5] 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, doi: 10.1016/S0967-070X(97)00011-5.

[6] 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, doi: 10.3141/2230-02.

[7] R. Aldred, T. Watson, R. Lovelace, and J. Woodcock, "Barriers to investing in cycling : Stakeholder views from England," *Transp. Res. Part A Policy Pract.*, 2018.

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

[9] D. Gettman, L. Pu, T. Sayed, and S. Shelgy, "Surrogate Safety Assessment Model and Validation : Final Report, Report No. FHWA--HRT-08-051," 2008.

[10] M. Bíl, M. Bílová, and I. Müller, "Critical factors in fatal collisions of adult cyclists with automobiles," *Accid. Anal. Prev.*, vol. 42, no. 6, pp. 1632–1636, 2010, doi: 10.1016/j.aap.2010.04.001.

[11] G. B. Rodgers, "Bicyclist deaths and fatality risk patterns," *Accid. Anal. Prev.*, vol. 27, no. 2, pp. 215–223, 1995, doi: 10.1016/0001-4575(94)00063-R.

[12] R. Aldred, J. Woodcock, and A. Goodman, "Does More Cycling Mean More Diversity in Cycling?," *Transp. Rev.*, vol. 36, no. 1, pp. 28–44, 2016, doi: 10.1080/01441647.2015.1014451.

[13] J. S. Mindell, D. Leslie, and M. Wardlaw, "Exposure-Based, 'Like-for-Like' Assessment of Road Safety by Travel Mode Using Routine Health Data," *PLoS One*, vol. 7, no. 12, pp. 1–10, 2012, doi: 10.1371/journal.pone.0050606.

[14] G. Welander, R. Ekman, L. Svanström, L. Schelp, and A. Karlsson, "Bicycle injuries in Western Sweden: a comparison between counties.," *Accid. Anal. Prev.*, vol. 31, no. 1–2, pp. 13–19, 1999, doi: 10.1016/S0001-4575(98)00040-2.

[15] F. E. Pedroso, F. Angriman, A. L. Bellows, and K. Taylor, "Bicycle use and cyclist safety following boston's bicycle infrastructure expansion, 2009-2012," *Am. J. Public Health*, vol. 106, no. 12, pp. 2171–2177, 2016, doi: 10.2105/AJPH.2016.303454.

[16] 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, doi: 10.1016/j.aap.2006.08.010.

[17] E. Hauer and A. S. Hakkert, "Extent and some implications of incomplete accident reportingNo Title," *Transp. Res. Rec.*, no. 1185, pp. 1–10, 1998.

[18] S. Meade and K. Stewart, "Vulnerable Road Users on Irish Roads , 2006-2012 : An Analysis of CT68 Data," in *STAR Conference, Glasgow*, May 2015.

[19] B. Sabey and H. Taylor, "The known risk we run: The Highway," Berkshire, 1980. [Online]. Available: https://trl.co.uk/sites/default/files/SR567.pdf.

[20] O. M. J. Carsten, M. R. Tight, M. T. Southwell, and B. Plows, "Urban accidents: why do they happen.," Basingstoke, 1989. [Online]. Available: https://trid.trb.org/view/349471.

[21] M. Hossain and Y. Muromachi, "A Framework for real-time crash prediction : Statistical Approach versus artificial intelligence," *Infrastruct. Plan. Rev. Japan Soc. Civ. Eng.*, vol. 26, no. 5, pp. 979–988, 2009.

[22] R. D. Connors, M. Maher, A. Wood, L. Mountain, and K. Ropkins, "Methodology for fitting and updating predictive accident models with trend," *Accid. Anal. Prev.*, vol. 56, pp. 82–94, 2013, doi: 10.1016/j.aap.2013.03.009.

[23] AASHTO, *Highway Safety Manual*, 1st ed. Washington. USA: American Association of State Highway and Transportation Officials, 2010.

[24] P. Greibe, "Accident prediction models for urban roads," *Accid. Prev. Prev.*, vol. 35, pp. 273–285, 2003.

[25] T. Jonsson, "Predictive models for accidents on urban links - A focus on vulnerable road users," Lund University, 2005.

[26] H. Peltola, "Evaluating road safety and safety effects using Empirical Bayes Method," in *4th IRTAD Conference*, 2009, pp. 307–314.

[27] M. Dougherty, "A review of neural networks applied to transport," *Transp. Res. Part C*, vol. 3, no. 4, pp. 247–260, 1995, doi: 10.1016/0968-090X(95)00009-8.

[28] S. Haykin, *Neural Networks, A Comprehensive Foundation*, Second Edi. Pearson Education (Singapore) Pte.Ltd, 2005.

[29] 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, doi: 10.1016/j.trc.2010.10.004.

[30] K. Gharehbaghi, "Artificial neural network for transportation infrastructure systems," *MATEC Web Conf.*, vol. 81, 2016, doi: 10.1051/matecconf/20168105001.

[31] K. Hajian-Tilaki, "Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation," *Casp. J. Intern. Med.*, vol. 4, no. 2, pp. 627–635, 2013.