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 distance traversed 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 elderly 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], 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 preference and re-
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 unable 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. PROPOSED 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 dashboard 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).](image1)

![Figure 2: Location and Boundaries of the study area](image2)
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 subgroup 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 applied as a data analytic method in supporting the nation. The neural network has been widely introduced 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.

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:

\[ O_a = \tanh(S_a) = \frac{e^{S_a} - e^{-S_a}}{e^{S_a} + e^{-S_a}} \]  

where \( O_a \) is the activation of the \( a \)th output neuron.

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

\[ O_a = \sigma(S_a) = \frac{e^{S_a}}{\sum_{k=1}^{m} e^{S_k}} \]  

where \( m \) 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.

\[ E = -\sum_{a=1}^{m} t_a \ln O_a \]  

where \( O_a \) 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

\[ \frac{\partial E}{\partial w_{ha}} = \sum_{a=1}^{m} (O_a - t_a) x_h w_{ha}(1 - x_h)x_h \]  

b) nodes between the output and hidden and layer

\[ \frac{\partial E}{\partial w_{hj}} = (O_a - t_a)x_h \]  

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

\[ \Delta w_{bh} = -\gamma \frac{\partial E}{\partial w_{ha}} \]  

\[ \Delta w_{bh+1} = w_{bh} + \Delta w_{bh} \]

\( x \) is the input variable, and \( \gamma \) 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

<table>
<thead>
<tr>
<th>Stopping and Memory Criterion</th>
<th>Steps (max) without a change in the error</th>
<th>999</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training (max) time</td>
<td>Training (max) epochs</td>
<td>999</td>
</tr>
<tr>
<td>Relative change in the training error (min)</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Relative change in the training error ratio (min)</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Cases to store in the memory (max)</td>
<td>999</td>
<td></td>
</tr>
<tr>
<td>Hidden layers</td>
<td>Total No. of Hidden Layers</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Elements in each layer</td>
<td>30</td>
</tr>
</tbody>
</table>

**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>
<thead>
<tr>
<th>Network Topology</th>
<th>Number of hidden layers</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Elements in each layer</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>Activation function between the hidden layers</td>
<td>Hyperbolic Tangent</td>
</tr>
<tr>
<td></td>
<td>Activation function between hidden and output layers</td>
<td>SoftMax</td>
</tr>
<tr>
<td></td>
<td>Error function</td>
<td>Cross-entropy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Training Type</th>
<th>Error function</th>
<th>Batch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Lambda</td>
<td>0.00000001</td>
<td></td>
</tr>
<tr>
<td>Initial Sigma</td>
<td>0.000001</td>
<td></td>
</tr>
<tr>
<td>Initial Centre</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Initial offset</td>
<td>40.001</td>
<td></td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Age</th>
<th>Frequency</th>
<th>Per cent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>17-24</td>
<td>537</td>
<td>16.2</td>
<td>16.2</td>
</tr>
<tr>
<td>25-34</td>
<td>494</td>
<td>14.9</td>
<td>31.1</td>
</tr>
<tr>
<td>35-44</td>
<td>347</td>
<td>10.4</td>
<td>41.5</td>
</tr>
<tr>
<td>45-54</td>
<td>251</td>
<td>7.5</td>
<td>49.0</td>
</tr>
<tr>
<td>55-64</td>
<td>115</td>
<td>3.5</td>
<td>52.5</td>
</tr>
<tr>
<td>Over 64</td>
<td>65</td>
<td>2.0</td>
<td>54.4</td>
</tr>
<tr>
<td>Under 17</td>
<td>1420</td>
<td>42.7</td>
<td>97.1</td>
</tr>
<tr>
<td>Unknown</td>
<td>96</td>
<td>2.9</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>3325</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Crash distribution across gender

<table>
<thead>
<tr>
<th>Gender</th>
<th>Collision Severity</th>
<th>Fatal</th>
<th>Serious</th>
<th>Slight</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>Number</td>
<td>1</td>
<td>90</td>
<td>278</td>
<td>369</td>
</tr>
<tr>
<td></td>
<td>Percentage</td>
<td>0.03%</td>
<td>2.71%</td>
<td>8.36%</td>
<td>11.10%</td>
</tr>
<tr>
<td>Male</td>
<td>Number</td>
<td>25</td>
<td>571</td>
<td>2360</td>
<td>2956</td>
</tr>
<tr>
<td></td>
<td>Percentage</td>
<td>0.75%</td>
<td>17.17%</td>
<td>70.98%</td>
<td>88.90%</td>
</tr>
</tbody>
</table>

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

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

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

<table>
<thead>
<tr>
<th>No.</th>
<th>Input Variable</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Spatial</td>
<td></td>
</tr>
<tr>
<td>a)</td>
<td>Month of Journey</td>
<td>Jan-Dec.</td>
</tr>
<tr>
<td>b)</td>
<td>Journey Day</td>
<td>Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday.</td>
</tr>
<tr>
<td>c)</td>
<td>Journey Weekday/Weekend</td>
<td>Weekday, Weekend.</td>
</tr>
<tr>
<td>d)</td>
<td>Journey Hour</td>
<td>0-23.</td>
</tr>
<tr>
<td>e)</td>
<td>Number of vehicles</td>
<td>1-5.</td>
</tr>
<tr>
<td>f)</td>
<td>Journey Purpose</td>
<td>Commuting, work trip, School Journey by Pupil, taking pupil to school, other, Unknown.</td>
</tr>
<tr>
<td>2.</td>
<td>Environmental</td>
<td></td>
</tr>
<tr>
<td>a)</td>
<td>Lighting conditions</td>
<td>Daylight /Darkness- No Street Lighting, Street Lighting Unknown, Street Lights present and lit, Street Lights present but unlit.</td>
</tr>
<tr>
<td>b)</td>
<td>Weather (Meteorological) conditions</td>
<td>Fine/Rain/Snow-with high winds, without high winds, fog, or Mist Hazard, Other.</td>
</tr>
<tr>
<td>c)</td>
<td>Road Surface Condition</td>
<td>Dry, Frost/ice, Wet/damp, Snow</td>
</tr>
<tr>
<td>3.</td>
<td>Infrastructure</td>
<td></td>
</tr>
<tr>
<td>a)</td>
<td>Road Type</td>
<td>Dual Carriageway, One-way street, Roundabout, single carriageway, slip road.</td>
</tr>
<tr>
<td>b)</td>
<td>Speed limit</td>
<td>20-70</td>
</tr>
<tr>
<td>c)</td>
<td>1st Road Class</td>
<td>A, B, C, E, U</td>
</tr>
<tr>
<td>d)</td>
<td>Road Hierarchy Level</td>
<td>0-4</td>
</tr>
<tr>
<td>e)</td>
<td>Road Hierarchy level and direction</td>
<td>-4 to 4</td>
</tr>
<tr>
<td>f)</td>
<td>Junction Detail</td>
<td>Crossroad, Mini Roundabout, Multiple Junction, Straight Road, Roundabout, Slip Road, T or Staggered, Private Drive</td>
</tr>
<tr>
<td>g)</td>
<td>Junction Control</td>
<td>No Control, Traffic Signal, Give way or uncontrolled, Stop sign</td>
</tr>
<tr>
<td>h)</td>
<td>2nd Road Class</td>
<td>A, B, C, E, U</td>
</tr>
<tr>
<td>i)</td>
<td>Vehicle Maneuver</td>
<td>Changing lanes, Going ahead, Moving off, Overtaking, Parked, Reversing, Slowing/stopping, Turning, U-turn, Waiting to go ahead, waiting to turn</td>
</tr>
<tr>
<td>j)</td>
<td>Vehicle Junction Location</td>
<td>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)</td>
</tr>
<tr>
<td>k)</td>
<td>Road Location of vehicle</td>
<td>Bus Lane, Busway, Cycle lane, cycleway, footpath, on layby or hard shoulder, main carriageway, tram/light rail track</td>
</tr>
<tr>
<td>l)</td>
<td>Skidding and Overturning</td>
<td>No skidding or overturning or jack-knifing, overturned, skidded, overturned, and skidded</td>
</tr>
<tr>
<td>Output Variable</td>
<td>Risk gender and Age Group</td>
<td></td>
</tr>
</tbody>
</table>

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**

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

<table>
<thead>
<tr>
<th>1. Spatial</th>
<th>Importance</th>
<th>Normalized Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Month of Journey</td>
<td>0.05</td>
<td>45%</td>
</tr>
<tr>
<td>b) Journey Day</td>
<td>0.05</td>
<td>42%</td>
</tr>
<tr>
<td>c) Journey Weekday/Weekend</td>
<td>0.02</td>
<td>20%</td>
</tr>
<tr>
<td>d) Journey Hour</td>
<td>0.06</td>
<td>55%</td>
</tr>
<tr>
<td>e) Number of vehicles</td>
<td>0.05</td>
<td>41%</td>
</tr>
<tr>
<td>f) Journey Purpose</td>
<td>0.11</td>
<td>100%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2. Environmental</th>
<th>Importance</th>
<th>Normalized Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Lighting conditions</td>
<td>0.05</td>
<td>48.2%</td>
</tr>
<tr>
<td>b) Lighting and road surface condition</td>
<td>0.06</td>
<td>50.9%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3. Infrastructure</th>
<th>Importance</th>
<th>Normalized Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Road Type</td>
<td>0.04</td>
<td>36.6%</td>
</tr>
<tr>
<td>b) Speed limit</td>
<td>0.05</td>
<td>41.1%</td>
</tr>
<tr>
<td>c) 1st Road Class</td>
<td>0.04</td>
<td>35.5%</td>
</tr>
<tr>
<td>d) Road Hierarchy level and direction</td>
<td>0.05</td>
<td>43.7%</td>
</tr>
<tr>
<td>e) Junction Detail</td>
<td>0.05</td>
<td>42.1%</td>
</tr>
<tr>
<td>f) Junction Control</td>
<td>0.04</td>
<td>31.9%</td>
</tr>
<tr>
<td>g) 2nd Road Class</td>
<td>0.04</td>
<td>37.4%</td>
</tr>
<tr>
<td>h) Vehicle Maneuver</td>
<td>0.05</td>
<td>46.3%</td>
</tr>
<tr>
<td>i) Road Location of vehicle</td>
<td>0.05</td>
<td>45.6%</td>
</tr>
<tr>
<td>j) Vehicle Junction Location</td>
<td>0.05</td>
<td>45.4%</td>
</tr>
<tr>
<td>k) Skidding and Overturning</td>
<td>0.03</td>
<td>29.1%</td>
</tr>
</tbody>
</table>

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