Empirical Modelling for exploring the factors contributing to disability severity from road traffic accidents in Thailand

Jaratsri Rungrattanaubol[1], **Anamai Na-udom**[2], and **Antony Harfield**[3], Non-members

ABSTRACT

This paper introduces a computer-based model for predicting the severity of injuries in road traffic accidents. Using accident data from surveys at hospitals in Thailand, standard data mining techniques were applied to train and test a multilayer perceptron neural network. The resulting neural network specification was loaded into an interactive environment called EDEN that enables further exploration of the computer-based model. Although the model can be used for the classification of accident data in terms of injury severity (in a similar way to other data mining tools), the EDEN tool enables deeper exploration of the underlying factors that might affect injury severity in road traffic accidents. The aim of this paper is to describe the development of the computer-based model and to demonstrate the potential of EDEN as an interactive tool for knowledge discovery.

Keywords: Disability, Artificial Neural Network, Interactive Environments, Data Mining

1. INTRODUCTION

Nowadays medical institutions are looking to find ways to reduce fatalities, disabilities and injuries caused by road traffic accidents, including the recommendation of precautionary measures and laws to prevent accidents. However, road traffic injury is still common, and it affects public health and quality of life, especially when the impact is death or disability. Many institutions have accumulatively recorded details of incidence of death or disability in order to find the causes of road traffic accidents, and the amount of collected data is very large. Many researchers have applied both statistics methods and data mining techniques to investigate the behaviour, characteristics and risks of road traffic accidents and their effect. Alcohol consumption is proposed to be one significant factor of accidents and fatality [1][2] and not-wearing seat belts and helmets is also a factor leading to severe injury [3].

In Thailand road traffic accidents are the second highest cause of death (the highest cause of death is cancer). In 1998 there were up to 7,986 deaths and in 2002 there were 13,438 deaths, increasing almost 70% within 4 years [4]. In this paper we are concerned with understanding the factors leading to death or disability resulting from road traffic accidents. The incidence of death and disability from road traffic injury has been surveyed and analysed by the Sirindhorn National Medical Rehabilitation Centre (SNMRC), Thailand, in 2006 [5]. In this dataset, there are two levels of disability, 'disability inclusion criteria' and 'disability non-inclusion criteria', as stated by the Rehabilitation for Disabled Person Act B.E. 2534 (1991) [6]. The 'disability inclusion criteria' level is considered more severe than the 'disability non-inclusion criteria' level. Persons diagnosed as 'disability inclusion criteria' are often unable to work and may be entitled to disability benefits. Persons classed as 'disability non-inclusion criteria' are less severe, they should be able to work and do not require as much support as the previous group. This paper focuses on using the dataset from the survey to construct a computer-based model that can predict the level of disability and that can be used to explore the factors that lead to disability following road traffic accidents.

Empirical Modelling (EM) is an extensive body of research developed by Beynon and Russ at the University of Warwick that focuses on computer-based interactive environments that are fundamentally different to traditional software construction (i.e. programming) [11]. EM is a user-centred approach to model building that focuses on three key concepts: observation, dependency and agency. The principles of EM hold that model building is essentially a sense-making activity in which the user is *observing* the situation, reasoning about the *dependency* (cause and effects) in the situation and exploring the actions (or *agency*) possible in the situation. EDEN is one of the EM tools that has been developed for building and exploring models based on these principles. In this paper, the authors have employed EM principles and used the EDEN tool to explore the factors contributing to injury in road traffic accidents.

Manuscript received on September 1, 2011 ; revised on January 31, 2012.

[1,3] The authors are with Department of Computer Science and IT,Faculty of Science, Naresuan University, Phitsanulok, Thailand., E-mail: jaratsrir@nu.ac.th and antonyh@nu.ac.th

[2] The author is with Department of Mathematics, Faculty of Science, Naresuan University, Phitsanulok, Thailand., E-mail: anamain@nu.ac.th

[3] Corresponding author.

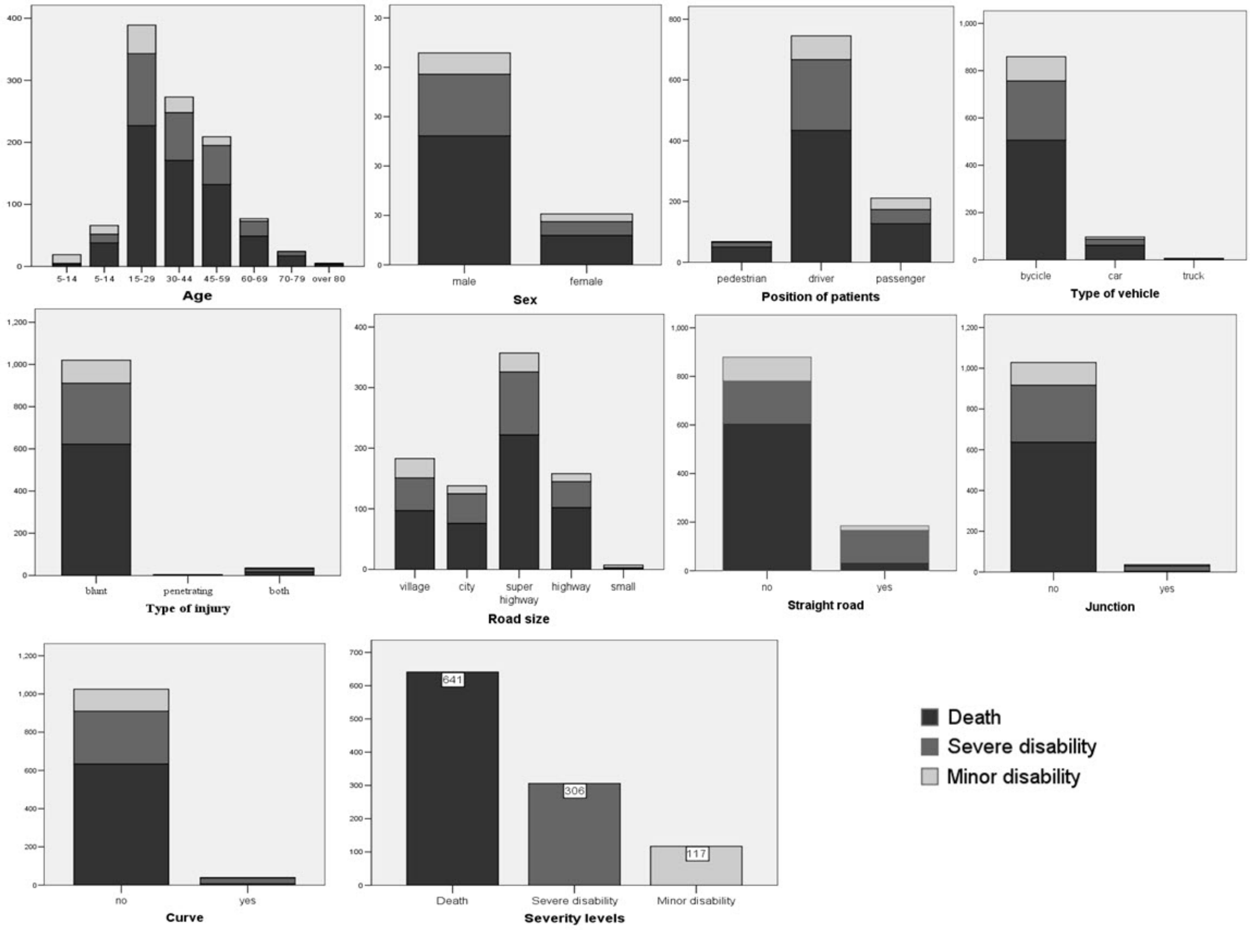

Fig.1: *Dataset input and output variable distribution.*

2. STATISTICAL DATA ANALYSIS

The dataset used in this research is a secondary injury surveillance data collected from 26 December 2005 to 25 June 2006 in 8 hospitals chosen from 28 sentinel sites (different provinces in various regions) around Thailand. The data is split into 2 groups, which are non-severely injured (i.e. not admitted) and severely injured (i.e. admitted). The surveillance period for non-severe cases is 3 weeks after patients left the hospital and for severe cases is 6 months. The total number of non-severely injured cases is 14,698 and the number of severely injured cases is 9,737. About 6.6% of the cases resulted in death and about 4.3% of the severely injured cases are diagnosed as disability. In order to predict the level of disability caused from road traffic injury, the cases that caused disability are selected for statistical data analysis.

The dataset consists of more than 150 variables, e.g. date and time of accident, sex, occupation, medical and alcohol taken, medical diagnosis, etc. In order to simplify the classification of the records, only statistically relevant variables are included in the model. Kendall's tau_b correlation coefficient (r) was used as the criterion for selecting the variables for the model. The input variables that are related to the output variables (level of disability, i.e. death, inclusion and non-inclusion disability) that are statistically significant at 0.30 level of significance are presented in Table 1. Only these nine variables are considered in the development of the model.

The input variables cover the personal characteristics of patients as well as the type of injury and the situation of the road accident.

Age is categorized into 8 categories (0-4, 5-14, 15-29, 30-44, 45-59, 60-69, 70-79, over 80). {1,2,3,,8}

Sex is 'male' or 'female'. 1,2

$Positionofpatient$ is 'pedestrian', 'driver', or 'passenger'. 1,2,3

$Typeofvehicle$ is 'bicycle/motorbike/three-wheel vehicle, 'car/pickup', or 'truck/lorry/bus/minibus'. {1,2,3}

$Typeofinjury$ is 'blunt', 'penetrating' or 'both'. {1,2,3}

$Roadsize$ is 'village road', 'city road', 'super highway', 'highway' or 'small road'. {1,2,3,4,5}

$Straightroad$ is 'no' or 'yes'. {0,1}

Junction is 'no' or 'yes'. {0,1}

$Curvedroad$ is 'no' or 'yes'. {0,1}

There is one output variable which is the severity of injury. This has three possible values:

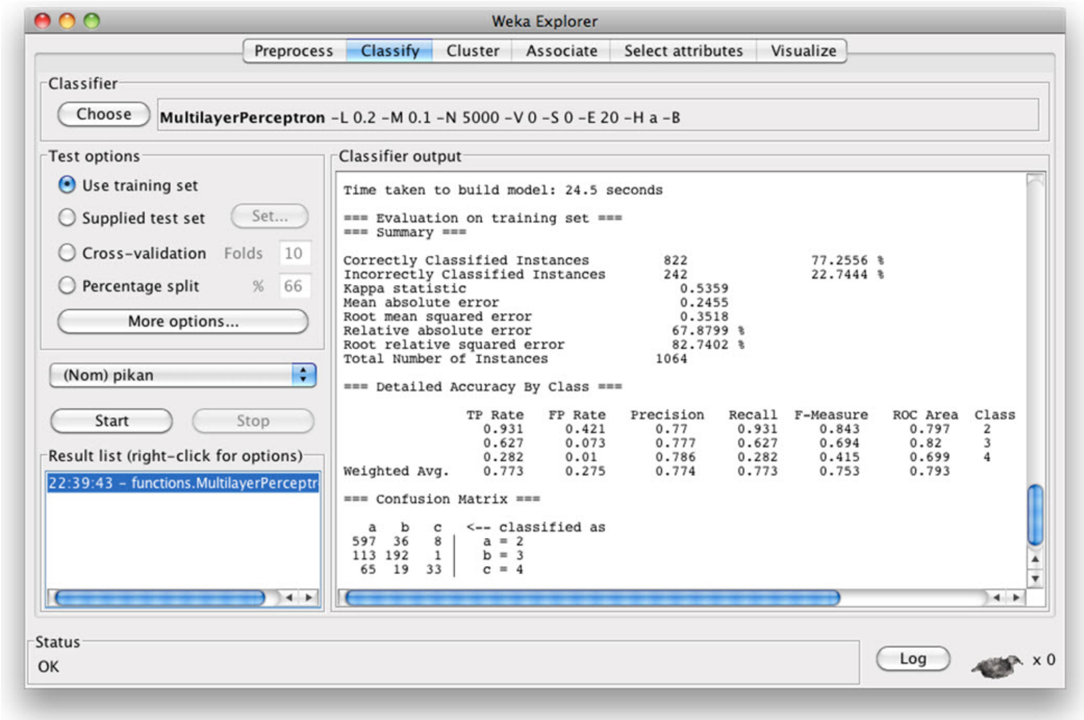

Fig.2: *Training and testing the Multilayer Perceptron in the Weka tool.*

1. Death
2. Disability inclusion criteria (Severe disability)
3. Disability non-inclusion criteria (Minor disability)

A disabled patient is defined as 'disability inclusion criteria' if he/she is a person with physical, intellectual or physical abnormality or impairment as categorized and prescribed in the Ministerial Regulation as stated in Rehabilitation for Disabled Person Act, B.E. 2534 (1991) [6]. The other type of disability is 'disability non-inclusion criteria' which is considered less severe (e.g. a person with unilateral blindness).

These nine input variables and one output variable are used to construct a neural network as described in the next section.

Table 1: *The selected significant variables.*

Variable Group	Variable name	r	P-value
Personal	Age(*age*)	-0.095	< 0.001
	Sex(*sex*)	0.035	0.240
	Position of patient(*injp*)	0.050	0.089
	Type of vehicle(*injt*)	-0.038	0.215
Injury	Type of injury (*tinj*)	0.064	0.030
Situation	Road size (*a31*)	-0.074	0.015
	Straight road (*a421*)	0.323	< 0.001
	Junction (*a43*)	0.164	< 0.001
	Curved road (*a45*)	0.131	< 0.001

The reason for only selecting data that resulted in death or disability is that the data set is quite large and there are a relatively large number of cases not resulting in serious injury. As early tests showed, a data set containing a large amount of less relevant data tends to create unnecessary noise for most data mining techniques. Therefore, in order to maintain a balanced data set with relatively even proportions of cases for each output, only the deaths and disabilities were selected. Figure 1 shows the distribution of the dataset in terms of the 3 possible values for the output variable.

3. USING NEURAL NETWORKS FOR DATA MINING

To construct a neural network for predicting the severity of injuries in road traffic accidents, the following steps were performed:

1. Data preparation and cleaning
2. Training the neural network using Weka
3. Presenting the neural network using Eden

The original data set was prepared in SPSS. There were a total of 9,739 records of accidents resulting in severe injuries, categorised as no disability, death, inclusion disability and non-inclusion disability. Some records had missing data and they were removed in order to simplify the training process. All records resulting in no disability were not included because early experiments showed that the large number of records for no disability tended to cause a bias. Therefore the final dataset contained only 1064 records for training and testing the neural network.

One continuous input variable is categorised, Age, which was divided into 8 distinct groups as mentioned in section 2.

The output variable *Severity* has 3 discrete values: {1,2,3}. The value 1 means the victim died as a result of the road traffic accident injuries. The value 2 means 'disability inclusion criteria' and the value 3 means 'disability non-inclusion criteria'. Of the 1064 records, 641 are categorised as 'death', 306 are categorised 'disability inclusion criteria' and 117 are categorised as 'disability non-inclusion criteria'. The distribution of the input and output variables is shown in Figure 1.

After cleaning the data, it was loaded into the Weka data mining tool [7]. Weka was configured to perform the training of a multilayer perceptron (MLP) (for background on multilayer perceptrons and neural networks see [16]). The MLP was chosen for analysis in this paper not because it is necessarily the best data mining classifier (although results in section 4.1 show that it is reasonably competitive) but because it is well-suited to the exploratory modelling as discussed in section 4.1. There were nine inputs (as chosen in section 2), and three outputs. Each output represents an injury severity level: 'out0' is death, 'out1' is severe disability and 'out2' is minor disability. Using Weka's suggested formula, the multilayer perceptron contained one hidden layer with six nodes. All of the nodes in the multilayer perceptron use a standard sigmoid function ($f(x) = (1+e^{-x})^{-1}$). The multilayer perceptron was trained using back-propagation, with a learning rate of 0.2, a momentum of 0.1, and a

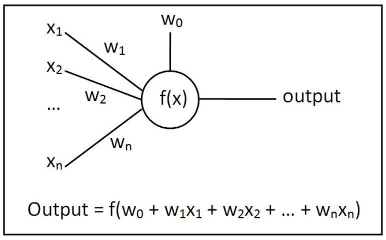

Fig.3: *Diagram of a perceptron.*

training time of 5000. The resulting model was tested with two different methods. Initially a cross-validation method was used to determine the accuracy of the MLP. This gave an accuracy value for comparison purposes (see results section 4.1). Following this, the neural network was re-trained using the full training set (i.e. all the data) to give the most accurate model for prediction purposes and further exploration. Figure 2 shows the configuration and output from Weka, for the training of the MLP and testing using the full training set.

The output from Weka gives the results of the training and the specification of the resulting neural network. The final step was to load the specification into EDEN (Engine for DEfinitive Notations) [8] for further exploration. EDEN is a general purpose interactive environment for creating and engaging with models or artefacts. An EDEN model consists of observables, dependencies and agency [9]. The observables in a model represent the current state, and often correspond to features that have meaning-such as the value of an input to the neural network corresponding to the age of an individual. The dependencies in a model represent the linkages between observables-for example, there is a dependency between the output of one perceptron and the input of another. The agency in a model corresponds to the external interference on state, usually by the user when they change, for example, the input values to the neural network or the configuration of the neural network itself. Models in EDEN therefore offer a highly flexible and open-ended environment for exploration.

A model was constructed in EDEN of the neural network trained using Weka. The constructed multilayer perceptron consists of a set of definitions, capturing the dependency in the model. As an example, the definition for one of the nodes in the hidden layer is:

```
node3 is sigmoid(-24.9 +
                 -17.8*age +
                 3.2*sex +
                 -25.5*injp +
                 4.9*injt +
                 45.2*tinj +
                 4.3*a31 +
                 3.7*a421 +
                 11.5*a43 +
                 -7.7*a45);
```

The above EDEN definition has a close association with the mathematical description of perceptron, as shown in Figure 3. The variables (age, sex, injp, etc) correspond to the inputs (x_1, x_2, x_3, etc) of the perceptron, and the values (-17.8, 3.2, -25.5, etc) correspond to the weights (w_1, w_2, w_3, etc) of the perceptron. There are similar definitions for all the components of the MLP, and the input for one perceptron is defined in terms of the output of perceptrons from a lower layer. In this respect, an EDEN model consists purely of a set of definitions, each of which corresponds closely to components of the MLP. In contrast, a procedural program representing an MLP consists of a sequence of steps that (when performed in a particular order) calculate the MLP output. However, the individual steps on their own do not have a direct correspondence to the components of the MLP. One of the key benefits of the EDEN environment is that constructing models using observables, dependency and agency allows a semantic correspondence between the software model and the real-world referent (in contrast to traditional

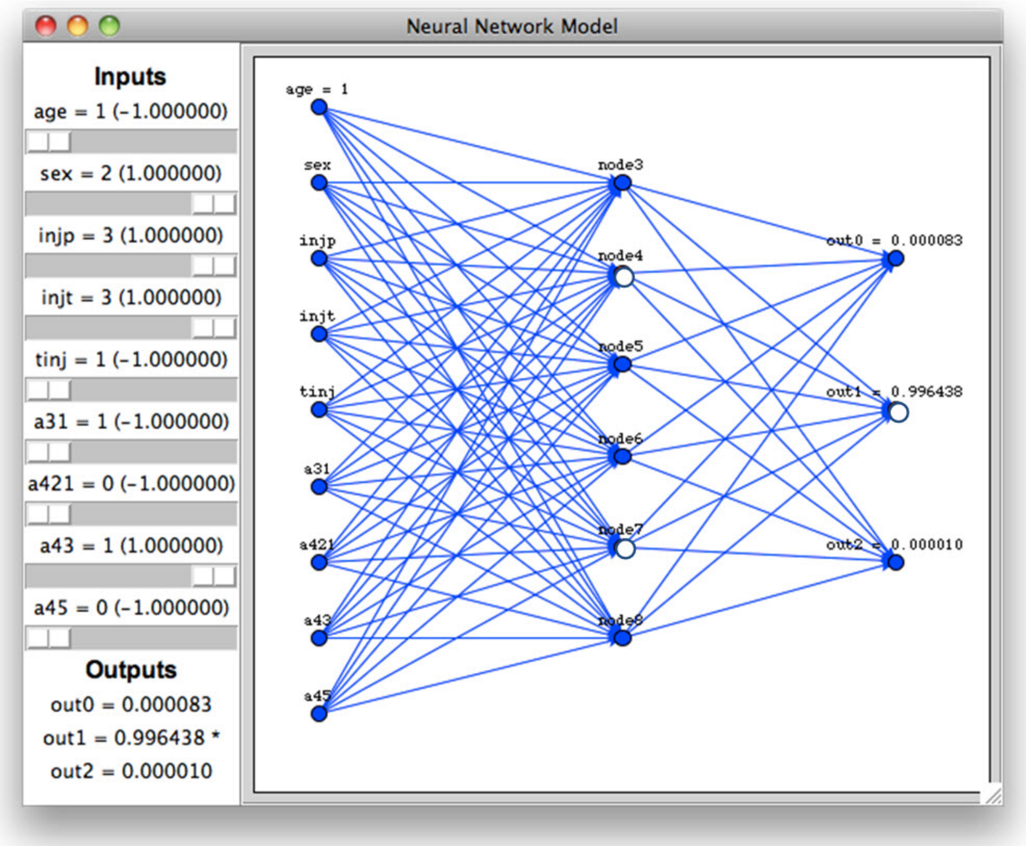

Fig.4: *The neural network model in EDEN.*

software programs [11]). This close, meaningful correspondence is irrelevant when the goal is speed and efficiency of computation (e.g. for data mining large datasets), but it is a desirable quality when concerned with experimentation and knowledge discovery, as will be shown in the next section. A complete discussion of the benefits of model-building in EDEN over procedural programming is beyond the scope of this paper (further evidence is presented in [11]).

On top of the basic model of the MLP (as trained using Weka), a user interface model was constructed to visual the neural network. A screenshot of the user interface (as modelled in EDEN) is shown in Figure 4. The user interface, like the model of the MLP, is a set of definitions. For example, each input/output variable on the left side of the screen is dependent on the underlying MLP. Any redefinition of the MLP could result in changes to the user interface. The flexibility for 'redefinition' is relevant to knowledge discovery because the effect of changes to the configuration of the neural network is immediately reflected in the user interface. This enables the user to explore a wide range of 'what-if' scenarios. For example, if the current configuration of inputs result in 'disability non-inclusion criteria' then a user can experiment by changing the value of an input in order to find out what variables cause the output to change to 'disability inclusion criteria' (more severe disability). Another user might be more interested in understanding the composition of the neural network and could change the connections between nodes in the MLP. As demonstrated in the next section, redefinitions of the interface allow insights into the model that point to the factors behind disability severity in road traffic accidents.

In previous work, EDEN was used for representing and exploring different types of neural network, including self-organising networks [12]. A notation for describing neural networks was developed using the principles of observables, dependency and agency. Such tools support knowledge discovery and learning in the area of neural networks. A closely related application area for EDEN is in decision support systems, as discussed by Beynon et al [13].

4. RESULTS AND ANALYSIS

There are two parts to the results: firstly we informally assess the accuracy of the model for predicting the level of disability from road traffic accidents; secondly we discuss the use of the model for understanding the factors that lead to disability from road traffic accidents.

4.1 Accuracy of the model

In order to judge the accuracy of the neural network, it is relevant to compare the achieved success rate with other data mining techniques. Weka was used to find the classification accuracy by measuring the success rate with the cross-validation technique over 10 folds. The results showed that the MLP produced a 73% success rate. A comparison with five other standard data mining classification techniques can be seen in Table 2. In this experiment, the MLP model is at least as accurate as the Random Tree model and the Radial Basis Function (RBF) model when tested using the cross-validation 10-fold technique. However, it is not as accurate as the Na?ve Bayes, Classification And Regression Tree (CART) and Support Vector Machines using Sequential Minimal Optimization (SVM) model, which each scored 75% accuracy.

The classification accuracy of the MLP can be improved by altering the number of hidden layers, and the number of nodes in the hidden layers. Table 3 shows the results of different combinations of hidden layers/nodes. The data shows that varying layers/nodes only affected the results by ~1%, and that the original configuration of 1 hidden layer with 6 nodes (as used in Table 2) is the best configuration for this dataset. Factors such as learning rate, momentum and number of iterations had an even lower affect on accuracy (<0.5%) and are therefore not presented here. This evidence suggests that 73% is the best accuracy that can be achieved with an MLP on the current dataset.

Although cross-validation accuracy appears low at 71-75%, it is important to note that the dataset contains records which have exactly the same inputs with conflicting outputs (as is normal for real-world datasets). This means it is impossible to achieve 100% accuracy. Further analysis of the dataset revealed that out of 1064 records, there are 841 records that have the same inputs, and 141 of these have conflicting outputs. In terms of accuracy, the maximum possible that could be achieved (irrespective of classification method) is 86%.

Table 2: *Comparison of classification techniques.*

Classification technique:	*Naive Bayes*	*CART*	*Rand Tree*	*RBF*	*SMO (SVM)*	*MLP*
Accuracy (Cross-validation)	75%	75%	71%	73%	75%	73%
Accuracy (full training)	74%	75%	80%	73%	75%	77%

Table 3: *Accuracy (%) of ANN with various hidden nodes and layers.*

No. of layers / nodes:	*1-hidden layer*			*2-hidden layers*			*3-hidden layers*	
	6	7	8	6:6	6:5	6:4	6:6:6	6:5:4
Accuracy (Cross-validation)	73	72	72	73	72	73	73	72

* learning rate=0.2, momentum=0.1, iteration no. = 5000

Table 4: *Confusion matrix.*

		Predicted output		
		1	2	3
Actual output	1	597	36	8
	2	113	192	1
	3	65	19	33

Another indication of the accuracy of the neural network can be observed from the confusion matrix. In Table 4, it can be observed that the neural network is somewhat biased towards an output of 1 (death), due to the training dataset containing a larger class size for this output (as shown previously in Figure 1). The network is 93% accurate for output 1, 62% accurate for output 2, and only 28% accurate for output 3. This demonstrates a could change the connections between nodes in the MLP. As demonstrated in the next section, redefinitions of the interface allow insights into the model that point to the factors behind disability severity in road traffic accidents.

In previous work, EDEN was used for representing and exploring different types of neural network, including self-organising networks [12]. A notation for describing neural networks was developed using the principles of observables, dependency and agency. Such tools support knowledge discovery and learning in the area of neural networks. A closely related application area for EDEN is in decision support systems, as discussed by Beynon et al [13].

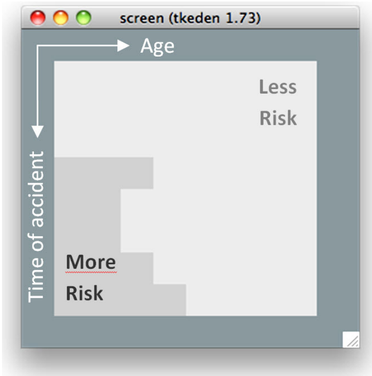

Fig.5: *Map of risk according to the MLP.*

4.2 Findings from the model

The second part of the results is concerned with what can be learnt about the factors that lead to death or disability in road traffic accidents. Using the EDEN environment, the factors affecting disability severity were explored by varying the inputs to the neural network. As discussed in the previous section, the visualisation can be redefined to give alternative views on the model. The example given here involves fixing some of the inputs and then observing the outcome of varying the remaining inputs. In one such exploratory exercise, from an earlier experiment [15] on a similar dataset concentrating on disability, the case is considered where a *pedestrian* is the victim of a *crash* where there was no *alcohol* or *drugs* or *safetydevices* involved, and there was no use of a *ventricleassistancedevice*. If these six variables are fixed then there are two remaining variables that can be explored: the age of the victim and the accident time of day. A new interface was defined on top of the existing model that shows the outputs of the model on an 8×8 grid (see Figure 5), representing all the possible *age* and *accidenttimeofday* values. The constructed visualisation shown in Figure 5 is shaded in a darker colour where the output of the MLP was 0, and a lighter colour where it was 1. According to this model, the group corresponding to the darker shading in the bottom left-hand corner is at risk of more severe disability from road traffic accidents in the specific case described above. This can be interpreted as younger people later in the day are more at risk. This is an example of one result that can be derived by redefining the interface to explore the model.

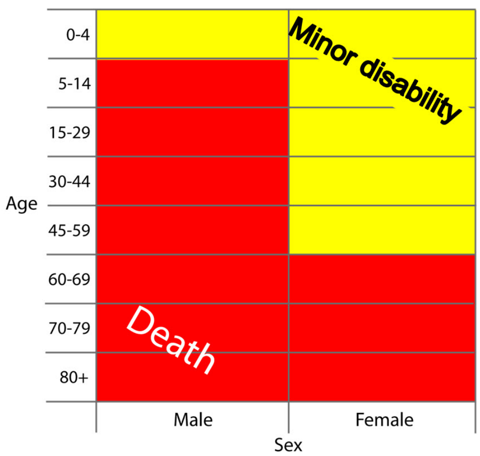

Fig.6: *The cases of motorcycle passengers involved in accidents on village roads.*

The same approach is used in this paper to highlight possible patterns in the neural network. The new results presented here are based on a more substantial dataset than the previous work [15].

The main argument in this paper is that EM principles and the EDEN tool offer powerful support for exploring and understanding the neural network and, in this case, how it predicts injury severity in road traffic accidents. In this section, we demonstrate the capability to answer typical 'what-if' questions that can point to specific results. In the simple case, consider setting the fixed variables for "a *passenger* on a *bicycle/motorcycle* with *blunt* injuries on a *villageroad*". The possible outcome of an accident in this situation can be visualised in terms of age and sex in Figure 6. Depending on the age and sex of the victim, there are two possible outcomes: death or minor disability. Figure 6 shows that elderly people are more susceptible to death, whereas younger people are more likely to escape death and receive only minor disabilities. In terms of sex, women are more likely to incur minor disabilities, even up to a later age. Whereas accidents involving men are more likely to result in death at any age.

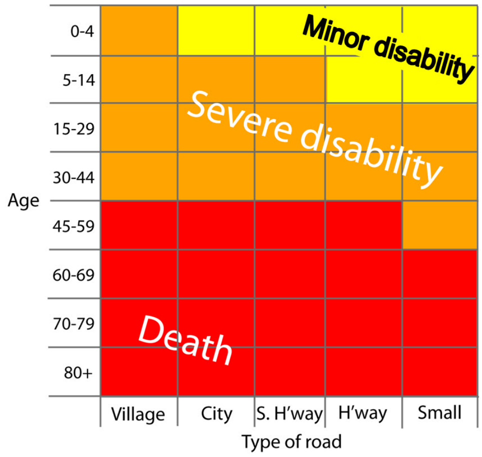

Fig.7: *The cases of male pedestrians involved in accidents with motorcycles (with blunt injuries, straight road, no junction, no curve).*

In other cases, the neural network produces only one output. For example, if you are the *driver* of a *truck/minibus/bus* involved in an accident at a *junction* with *both* blunt/penetrating injuries, then it is most likely that you will be in the severe disability category - no matter what *age* or *sex* you are. There are many conditions in which the only output is 'death'. One of the weaknesses of the neural network is that it is biased towards classifying the most common output (death), as described in the previous section and as shown by the confusion matrix in Table 3. By exploring the model in this way, the EDEN tool enables an expert to reveal the weakness in the model and the dataset.

Other cases show more variation in output. For example, in Figure 7 the two chosen variables are age and type of road. The selected conditions were male pedestrians involved in motorcycle accidents (on straight roads with no junctions or curves, resulting

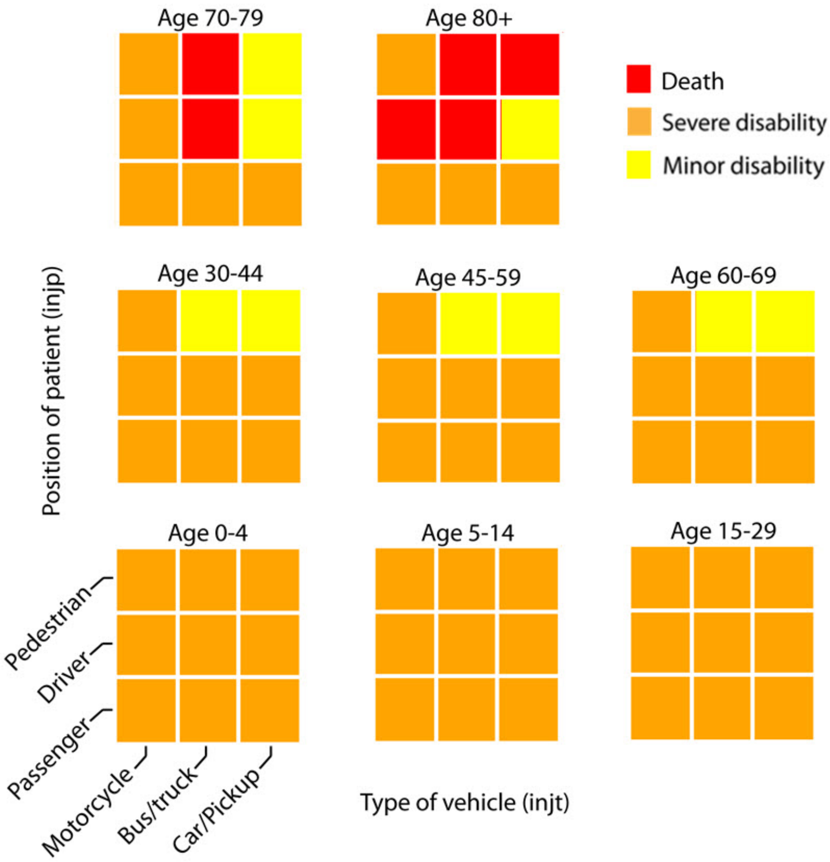

Fig. 8: *The cases of male pedestrians involved in accidents with motorcycles (with blunt injuries, straight road, no junction, no curve).*

in blunt injuries). In this case, it can be seen that elderly people are more at risk of death, and younger people are more likely to suffer disability. Very young people, 0-14 years, are more likely to only receive non-severe disabilities. It also shows a slight angle on the line between non-severe and severe disability, which could be interpreted that village roads are more dangerous (in terms of more deaths and severe disabilities resulting from accidents).

The two dimensional diagrams produced using the EDEN tool highlights specific cases from which results can be drawn. This method was extended further to examine the variation of 3 parameters. Consider the case where we are interested in female victims of all age categories, across all types of vehicle and all positions of the victim. This can be visualised using 8 screens shots organised in a grid, as shown in Figure 8. The lower part of the diagram shows that young girls/women (0-29 years old) are less at risk of death, and more likely to receive minor disability, from accidents in these conditions. As the age increases, the outcome of accidents is more unpredictable. The middle age group (30-69 years old) have some chance of severe disability, especially in the case where the victim is a pedestrian. In the upper part of the diagram, the results show that the elderly group (70+ years old) have more chance of death or severe disability (and less chance of minor disability).

Alternatively, Figure 8 could reveal flaws in the model. Considering the elderly group (70+ years old) in the upper part of the diagram, some might question the result that being a passenger is safer than being the driver. Furthermore, the results seem to suggest that a motorcycle is a safer form of transport than car or bus. These results point to errors in the model, possibly due to a lack of examples in the training data set for those conditions.

The above discussion demonstrates some results that are relevant to the field of medical science, and it is possible to draw conclusions from this work regarding what factors affect the outcome of road traffic accidents in Thailand. However the most important result from the computer science perspective is the method by which the insights were gained. The main aim of this paper is not to provide a complete analysis of the causes of disability resulting from road traffic accidents, the purpose here is to demonstrate the capability of the EDEN tool for exploration and experimentation to uncover new results, and to highlight possible errors in the results.

5. CONCLUSION

This paper has shown that a neural network is an appropriate data mining solution for the purpose of predicting disability severity from road traffic accidents in Thailand. For the selected dataset, a MLP neural network was shown to be appropriate as a classification technique. An accuracy of 77% was achieved, which is at least as accurate of as other competitive data mining classification techniques. In a similar study [10], different types of neural network were used to build models of the causes of road traffic accidents in Florida. Using a MLP they were able to predict injury severity based on a similar selection and number of input variables. The MLP also proved to be the most accurate type of neural network out of the methods used in their study. As is typical of real-world datasets, the data in our study is highly imbalanced which caused bias in the results. Attempts at sampling caused a dramatic loss of accuracy, and therefore further work is needed in this area. It is acknowledged that increased accuracy might prove difficult due to the imbalanced nature of real-world medical datasets as found here and in related work [10,14].

As a consequence of the current study, the resulting neural network can be used as an example of how to predict injury severity in terms of disability and death. By fixing key variables and experimenting with a subset of the variables, we can observe the factors that contribute to disability severity, as modelled by the neural network trained with the survey from Sirindhorn National Medical Rehabilitation Centre in Thailand. This paper only goes as far as highlighting the potential for understanding the factors behind accidents. Further work is needed, together with larger, more balanced datasets and better trained neural networks, in order to draw clear inferences regarding the factors that contribute to road accidents.

The final contribution of this paper has been to demonstrate the one potential use of the EDEN environment for constructing and exploring a computer-based model that enables knowledge discovery in the real-world domain of road traffic accidents. Previous work has demonstrated the potential for interactive environments such as EDEN to be utilised for data mining and knowledge discovery-the current study contributes another example of the need for software that encourages exploration and experimentation in the area of intelligent systems. Data mining is generally concerned with rigorously applying algorithms that can give a definite classification for a particular set of inputs. Therefore the computer acts as an independent decision maker, with little opportunity for the user/expert to examine the reasoning behind a decision. This paper introduces an alternative approach where the expert works 'with' the computer to explore and question the reasoning behind the classification. In this alternative approach, the computer is not an independent decision maker, but a tool for enhancing the human reasoning process.

6. ACKNOWLEDGEMENT

The authors would like to thank the Thailand Research Fund (TRF) for providing research funding and support for this project. The authors would also like to acknowledge Sirindhorn National Medical Rehabilitation Centre for providing the data set for this research.

References

[1] S. Zajac, J. Ivan, "Factors influencing injury severity of motor vehicle-crossing pedestrian crashes in rural Connecticut," *Accident Analysis and Prevention 2003* , 35(3), pp. 369–379.

[2] M. Keall, W. Frith, T. Patterson, "The influence of alcohol, age and number of passengers on the night-time risk of driver fatal injury in New Zealand," *Accident Analysis and Prevention 2004*, 36(1), pp. 49–61.

[3] F. Valent, F. Schiava, C. Savonitto, T. Gallo, S. Brusaferro, F. Barbone, "Risk factors for fatal road accidents in Udine, Italy," *Accident Analysis and Prevention 2002*, 34(1)71-84.

[4] Thai Health Promotion Foundation: The sustainability of well-being for Thai people. "Annual report 2003 - The Thailand Research Fund (TRF)," Bangkok, 2004.

[5] D. Suvaphan et al, "Incidence of Disability and Impact from Road Traffic Injury, 2006," *Sirindhorn National Medical Rehabilitation Centre (SNMRC)*, 2006.

[6] Rehabilitation for Disabled Person Act, B.E. 2534 (1991). URL: http://thailaws.com/law/t_laws/tlaw0245.pdf

[7] Weka 3. Data Mining Software in Java. URL: http://www.cs.waikato.ac.nz/ml/weka/

[8] Empirical Modelling Research Group. EDEN. URL: http://www2.warwick.ac.uk/fac/sci/dcs/research/em/

[9] J. Rungrattanaubol. A treatise on Modelling with definitive scripts. PhD thesis, Department of Computer Science, University of Warwick, April 2002.

[10] M.A. Abdel-Aty1, H.T. Abdelwahab. "Predicting Injury Severity Levels in Traffic Crashes: A Modeling Comparison," *Journal of Transportation Engineering*, Vol. 130, No. 2, March 1, 2004, pp.204–210.

[11] W.M. Beynon, R.C. Boyatt, S.B.Russ. "Rethinking Programming," In *Proceedings IEEE Third International Conference on Information Technology: New Generations (ITNG 2006)*, April 10-12, 2006, Las Vegas, pp.149–154.

[12] Anonymous. "Neural Networks and Notations," In The Second Warwick EM Bulletin (WEB-

EM-2). URL:
http://www2.warwick.ac.uk/fac/sci/dcs/research/em/publications/web-em/02/neural.pdf
[13] W.M. Beynon, S. Rasmequan, S. Russ. "A New Paradigm for Computer-Based Decision Support," *Decision Support Systems*, 2002, Vol 33, pp.127-142.
[14] N.V. Chawla, "Data Mining for Imbalanced Datasets: An Overview," Data Mining and Knowledge Discovery Handbook, 2010, pp.875-886.
[15] J. Rungrattanaubol, A. Na-udom, A. Harfield. "An exploratory neural network model for predicting disability severity from road traffic accidents in Thailand". In *3rd International Conference on Knowledge and Smart Technologies (KST-2011)*, July 16-17, 2011. Burapha University, Chonburi, Thailand.
[16] S. Haykin. Neural Networks: a Comprehensive Foundation, Macmillan, 1994.

Jaratsri Rungrattanaubol Ph.D. assistant professor, received her Ph.D. in Computer Science from Warwick University, England, in 2002. Currently she has worked as a lecturer and researcher at Dept. of Computer Science and IT, Faculty of Science, Naresuan University, Thailand for over 10 years. Her research area includes data mining, evolutionary search algorithms, computer simulated experiments and empirical modelling.

Anamai Na-udom Ph.D. assistant professor, graduated in Applied Statistics from Curtin University, Perth, Australia in 2006. She has been a statistics lecturer and researcher for over 10 years in Dept. of Mathematics, Faculty of Science, Naresuan University, Thailand. Her research interest includes statistical modelling, experimental design, evolutionary search algorithms and statistics in data mining techniques.

Antony Harfield received his Ph.D. in Computer Science from the University of Warwick, UK, where he developed novel approaches to educational technology using Empirical Modelling. Antony has worked as a researcher in academia and a technology consultant in industry. Currently he is a lecturer at Naresuan University, where he specialises in mobile technology.