

MEISAM SIAMIDOUDARAN, Ph.D.^{1,3}
(Corresponding author)
E-mail: meisam.siamidoudaran@emu.edu.tr

ERSUN ISCIOGLU, Ph.D.²
E-mail: ersun.iscioglu@emu.edu.tr

¹ Eastern Mediterranean University
Department of Civil Engineering,
99628, Famagusta, Cyprus

² Eastern Mediterranean University, Department
of Computer Education and Instructional Technology
99628, Famagusta, Cyprus

³ Voronezh State Technical University, Faculty of Economics,
Management, and Faculty Information Technology
Ulitsa 20-Letiya Oktyabrya, 84, 394006 Voronezh, Russia

Safety and Security in Traffic
Preliminary Communication
Submitted: 9 July 2018
Accepted: 29 Aug. 2019

INJURY SEVERITY PREDICTION OF TRAFFIC COLLISION BY APPLYING A SERIES OF NEURAL NETWORKS: THE CITY OF LONDON CASE STUDY

ABSTRACT

This paper focuses on predicting injury severity of a driver or rider by applying multi-layer perceptron (MLP), support vector machine (SVM), and a hybrid MLP-SVM method. By correlating the injury severity results and the influences that support their creation, this study was able to determine the key influences affecting the injury severity. The result indicated that the vehicle type, vehicle manoeuvre, lack of necessary crossing facilities for cyclists, 1st point of impact, and junction actions had a greater effect on the likelihood of injury severity. Following this indication, by maximising the prediction accuracies, a comparison between the models was made through exerting the most sensitive predictors in order to evaluate the models' performance against each other. The outcomes specified that the proposed hybrid model achieved a significant improvement in terms of prediction accuracy compared with other models.

KEY WORDS

road safety; traffic crash; injury severity prediction; contributory factors;

1. INTRODUCTION

Due to the dramatic increase in irreparable results of traffic collisions, road safety has become a worldwide concern in the transportation industry, and also in the developed countries such as Great Britain. The Britain's vision concerning road safety is to promptly mitigate the high risk of any issues and ensure that the UK lasts as an effective leader in road safety across the world. A variety of factors such as crash circumstances, characteristics of roadways, environment, atmosphere, road user behaviour, vehicle characteristics etc. can be highly effective in accidents.

The injury severity prediction model in traffic accidents is an appropriate procedure to better understand the relationship between the nature of the injury and the related contributory factors. As a result, several contributing factors can be identified in order to eliminate or mitigate the injuries. These factors alone can be significant, or when combined with each other they can contribute to the injury severity in traffic crashes, but in general, the reality is that the crashes are almost always caused by multiple correlated factors due to their complex interactions. However, there are still more contributory factors related to the injury severities that do not exist for road safety studies while the factors can affect the injuries [1]. Therefore, this study attempts to predict the injury severity of the driver or rider using STATS19 road safety data, which provides large numbers of subdivision for circumstances of personal injury variables [2]. In addition, because of the solid association between the related variables, using more input variables leads to getting better prediction accuracy [3].

2. LITERATURE REVIEW

This literature review is focused on various applications of driver injury severity prediction. For instance, in the past decades, typical statistical models were extremely common prediction models. Although the statistical models have the ability to detect a wide range of influencing factors, nowadays, the mass of complex data on accidents still makes it complicated to better understand the interactions of the non-linear injury severity-related factors while using these methods efficiently. In addition, a weak performance in using numerous separable variables as well as the variables

with a large number of categorical factors is another disadvantage [4]. Therefore, a number of researchers have cautioned of the linearity and some distributions of error terms by the statistical model. Accordingly, the applied statistical models fail when dealing with complex and extremely non-linear datasets [4-8].

To overcome the weakness of the traditional models, many studies have also developed artificial neural network (ANN) models for numerous problems of road safety due to their advantage of dealing with a mass of noisy data as well as superior predictive ability. The ANN models have proven to be quite a useful tool in predicting multiple variables and can establish better relationship between accidents and the related factors. As a result, they achieve a good model fit accompanied by prediction accuracy in comparison to the statistical models [1, 3, 7-12].

A commonly used network in these systems is the multi-layer perceptron (MLP) neural network. By drawing on two-vehicle accidents, Abdelwahab and Abdel-Aty [13] categorised traffic accident severity into three different categories according to injury types. The two researchers applied MLP and fuzzy Adaptive Resonance Theory (ART) with ordered logit models in their study to find the connection between the severity of the driver's injuries and other crash factors. These findings indicated that MLP classified the variables better than other models. Three years later, they compared the previous research results and found that MLP had a much more superior performance than fuzzy ART [14]. Additionally, Delen et al. [15] attempted to find the most important predictors of injury severity using a series of binary MLP neural networks. However, exploiting more injury levels along with the findings of the prediction did not offer any better solutions than other earlier studies [13, 14].

Literature relating to the SVM models which is also a quite commonly used type of the prediction was reviewed [8, 10, 11, 16]. For example, Li et al. [10] used this model for the prediction of injury severity and the obtained outcomes were compared with those obtained through the ordered probit model. The comparison results indicated that the SVM model produced results with a better prediction accuracy in recognising the significant predictors. Another comparison was applied by Li et al. [8] between SVM and NB regression and the results showed that the prediction accuracy resulting from SVM is superior to the NB model. In recent years, Yu and Abdel-Aty [16] have developed accident injury severity analysis models using SVM, random parameter logit model, and fixed parameter. After applying the models, the researchers found that both SVM and the random parameter models provided greater prediction accuracy resulting in a comparison between the models. In addition, the results showed that it is important to consider the potential non-linearity and individual heterogeneity while analysing

the factors. Most recently, Chen et al. [17] used the SVM models to examine injury severity classes. The outcomes of the predictions showed that the performance of the SVM models was satisfactory and the polynomial kernel outperformed the Gaussian RBF kernel. Another recent comparison has been carried out by Iranitalaba and Khattakb [18] by applying several statistical and machine-learning models. The outcome of the comparison indicated that the nearest neighbour algorithm attained the greater overall performance. Random forests and SVM models had the following two satisfactory performances and Multinomial Logit was the frailest model.

Another related comparison was made by Xie, Lord, and Zhang [1] and the outcome displayed that the Bayesian and back-propagation methods achieved superior predictive accuracy by comparison with the NB regression models.

Furthermore, artificial intelligence and non-parametric methods have been developed to resolve the limitation of statistical methods. For example, Chang and Wang used the classification and regression trees to determine the relationship among several classes of accidents and reasons influencing the crash severity [19].

Briefly, the considered previous studies indicated that the traditional methods were more commonly used, but the ANN models had higher prediction accuracy, mostly the MLP and SVM models. Therefore, using traditional models has been ignored in this study. Particular attention was placed on the modelling of the crash predictions. However, injury-related results of the predictions were not a common key focus with the aim of categorising the influences. Likewise, responding to the lack of the contributory factors-related study, an attempt was made to predict the injury severity levels by using various ANN algorithms along with various factors that contribute to the risk of injuries. Consequently, the MLP and SVM models with varying levels of achievement in previous studies were addressed as an identification technique for sensitive predictors in this study. Within this framework, the models were applied with the traffic crash data for the entire backbone of the road networks for the city of London. Also, the model forecasts the injury severity into either of the following classes: 1 – fatal injury, 2 – serious injury, 3 – slight injury, and 4 – only damage to property. According to the obtained results of the used prediction models and the potential of the injury severity analysis, the relationships between non-linear injury risks and the influencing variables were apprehended. Accordingly, the most significant predictors were ordered and considered as the main contributing factors to the injuries. Furthermore, the second stage of the predictions has been carried out using the more sensitive predictors through MLP, SVM, and a proposed hybrid MLP-SVM model. For the first time ever,

the hybrid MLP-SVM model has been used in an injury severity-related study with the objective of maximising the forecast accuracy. Accordingly, a comparison between the models was employed. The findings of this study can play a vital role in helping casualty reduction and prevention targets along with handling numerous road safety issues. It is hoped that these algorithms and models could shed some light on the potential of remedial measures and techniques to prevent such accidents.

3. MATERIALS AND MODELS

3.1 Data description

As a result of there being a strong relationship between the injury severity and the relevant variables, using fewer than seven input variables leads to getting worse prediction accuracy [3]. Therefore, this research attempted to import a larger mass of input data so as to reduce the accuracy error. The data used in this study were obtained from the STATS19 road accident data, excluding the crashes involving pedestrians that occurred between 2010 and 2015 in the city of London total road network. Consequently, the data indicated that the overall sum of 3,502 drivers/riders was involved in the reported accidents. The dataset provides a collection of the traffic collisions, which resulted in personal injury or death. The data cover the police recorded accidents at the roadside or those that were reported by a member of the public to a police station within 30 days of the collision. This set of data refers to a relational database and an underlying structure of the data that are fundamentally hierarchical. The

data content consists of three major variables including accident circumstances, vehicle, and casualty records. The variables detail all the explanatory factors involved in collisions as shown in *Table 1*. Moreover, it is referred to the UK Department for Transport (DFT) [2] for a detailed description of the data used in this study.

3.2 Models

3.2.1 MLP neural network

MLP is a commonly applied neural network that can be used to categorize various road safety problems. Some prior studies validated the advanced potential of using the MLP classifier in the predicting injury severity [13-15]. The MLP neural network is composed of three parts and in each part, some neurons exist. Every neuron in the input layer corresponds to the input indices. Also, in the hidden layer, the number of neurons was obtained based on trial-and-error to reach the optimum value, and finally the identified neurons for the output layer are considered based on the output indices. Like human neural networks, in MLP, there are connections between each layer of nodes with each node having a different weight. The MLP network requires an accurate algorithm to adjust its weights and bias terms by using good data. For this purpose, the Levenberg-Marquardt backpropagation algorithm was used since it works best with the type of data used in this research. The algorithm used in this study was developed through the combination of the gradient descent and Gauss-Newton algorithms that have been capable of predicting performance by using

Table 1 – Export variables of STATS19 data

Major explanatory variables (Model input parameters)	
Accident circumstances	Vehicle involved
Road type	Vehicle type
Road surface condition	Vehicle manoeuvre
Junction location	Vehicle propulsion
Junction control* & Junction detail*	Gender of driver/rider
Weather condition	Age of driver/rider
Lighting condition	Urban or rural area
Speed limit	Engine capacity
Number of vehicles	Junction location*
Number of casualties	1 st point of impact
Time, year, month, and day	Age of vehicle (manufacture)
1 st road class & 2 nd road class*	Driver index of multiple deprivation (IMD)
Casualty (model output parameter)	Injury severity classes of driver / rider
Pedestrian crossing facilities**	Journey purpose of driver

* This factor is used while the accident is at intersection, otherwise the value is void for the prediction model.

** This factor is applied for controlling pedestrians and cyclists crossing [2].

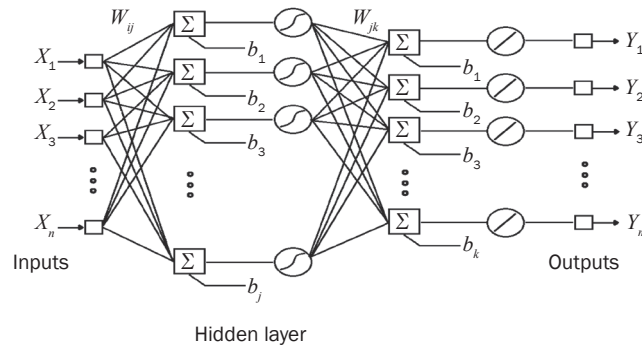


Figure 1 – Architecture of MLP model for injury severity prediction

various data. The Levenberg–Marquardt backpropagation algorithm was also referred to as the least-squares damped method and it decreases the function by employing a numerical solution. It involves two important steps: the propagation and weight update. In order to perform these steps, the algorithm uses the information from the front-propagation to calculate the neuron outputs for each layer. Later, by considering the training pattern target, each layer weight gradient was calculated through the difference between each layer target and output, and then finally each layer weight was updated. The amount of each neuron in the hidden layer is computed by using Equation 1.

$$P_j = f\left(\sum_{i=1}^A x_i^T \cdot w_{ij} + b_j\right) \tag{1}$$

where f represents the activation function for the hidden layer; A is the input neuron of each layer which is the number of traffic accidents in this study; x_i corresponds to the input layer neuron values; w_{ij} stands for weight between i -th input layer neuron and j -th hidden layer neuron; and b_j represents the amount of bias for the hidden layer. Furthermore, the value of output layer for every neuron is calculated by Equation 2.

$$y_k = g\left(\sum_{j=1}^B p_j^T \cdot w_{jk} + b_k\right) \tag{2}$$

If g arises as the linear transmission function, B represents the value for each neuron in the hidden layer; p_j corresponds to the hidden layer neuron values; w_{jk} is the amount of weight between j -th hidden layer neuron and k -th output layer neuron; and b_k symbolizes the amount of bias for the output layer [13, 23].

3.2.2 SVM neural network

The SVM model is a relatively new technique of machine learning which is used to examine the classification problems, and has been applied for the classification of collision injury severity [8, 10, 11, 16, 17]. This technique can identify hyperplanes between the groups or classes in a certain dimension space of available inputs. That is, the larger the margins between the locations of different groups or classes,

the better the generalisation of error classifier. For instance, take a sample of training data as $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, where each dataset possesses the inputs $(x_i \in R^d)$ also a group or class name as $(y_i \in \{+1, -1\})$ (see Figure 2). Furthermore, for the space of two dimensions, a line usually separates the classes in the middle of the margin which is referred to as the discriminator. However, for multi-dimensional spaces, a hyperplane usually separates the classes. If the distance taken between each dataset and the discriminator is one, then two hyperplanes are needed in parallel and the function of the classifier is calculated through Equation 3.

$$\begin{cases} w^T \cdot x_i + b \geq 1, & \text{if } y_i = 1 \quad i = 1, 2, \dots, n \\ w^T \cdot x_i + b \leq -1, & \text{if } y_i = -1 \quad i = 1, 2, \dots, n \end{cases} \tag{3}$$

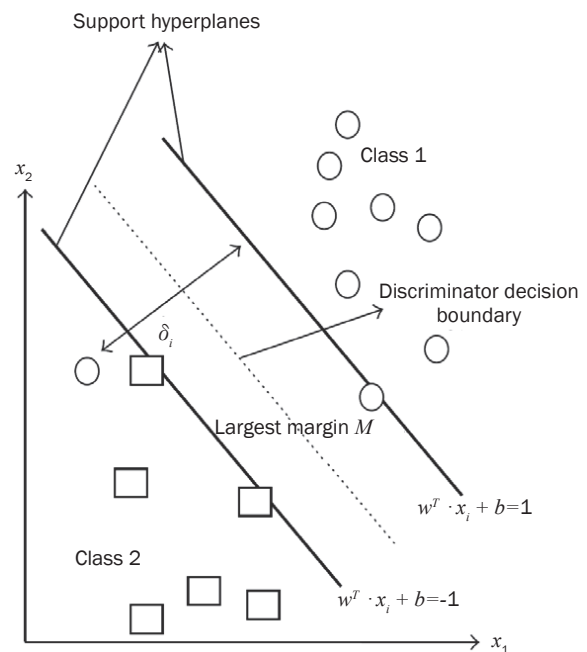


Figure 2 – A sample of linear soft margin SVM

If a specific discriminator needs to be obtained, the margins between the classes or groups have to be the highest. Therefore, by considering the distance between the support hyperplanes as M , the best

possible margin (M) will be calculated as $\frac{1}{2}\|w\|^2$. However, in nature, the data are not usually separate and they rather occur in multiple datasets; thus, as a result, a hyperplane or separator is obtained according to the least number of errors. Hence, the members of each class are determined by considering the distance from its own class (δ) borderline. This is the underlying strategy of the soft margin of the SVM Model. Therefore, non-negative parameters (δ_i) can be considered as inactive parameters based on (s.t.) $\delta_i \geq 0$. As a result, the primal problem is as follows:

$$\text{Minimise } \frac{1}{2}w^T \cdot w$$

s.t.:

$$y_i(w^T \cdot x_i + b) \geq 1 - \delta_i \quad i = 1, 2, \dots, n \tag{4}$$

$$\delta_i \geq 0$$

therefore,

$$L_p = \frac{1}{2}w^T \cdot w - \sum_{i=1}^n [y_i(w^T \cdot x_i + b) - 1 + \delta_i] \quad i = 1, 2, \dots, n \tag{5}$$

The primal problem is to do with a quadratic program which needs some efforts to be solved since input vector parameters are not the only parameters to be considered; that is, the problem relatively needs to take into account a number of other parameters as well. As a result, through the application of the Lagrange method, the primal form of the equation takes a dual form. As the Lagrange coefficients (α_i, μ_i) are required to be non-negative real factors, the fifth equation is as follows:

$$L_p = \frac{1}{2}w^T \cdot w + C \sum_{i=1}^n \delta_i - \sum_{i=1}^n \alpha_i [y_i(w^T \cdot x_i + b) - 1 + \delta_i] - \sum_{i=1}^n \mu_i \cdot \delta_i \quad \alpha_i, \mu_i \geq 0 \tag{6}$$

where C is considered to be the penalty factor and L_p is a saddle point; therefore, the lowest amount must be used by considering variables w, b and δ and the highest amount needs to be set by considering the Lagrange multipliers (α_i, μ_i). In order to shift the primary issue to a maximised issue, the partial derivative of w, b and δ needs to be taken into account. By doing so, the following dual problem is acquired:

$$L_D = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \cdot \alpha_j \cdot y_i \cdot y_j \cdot x_i^T \cdot x_j + \sum_{i=1}^n \alpha_i \tag{7}$$

After the process of formulas and calculations, the optimal decision function for hyperplane is computed as below:

$$y_i = \text{sign}(w^T \cdot x_i + b) \tag{8}$$

It is very difficult to select the best hyperplane to separate the non-linear data. To rectify the problem, the Hilbert-Schmidt theorem has been employed to change the d -dimensional input vector x to a higher-dimensional feature vector through the application of a

multi-dimensional vector function ($\emptyset:R^d \rightarrow R^N$), for which the obtained decision function of the optimal hyperplane appears as below:

$$y = \text{sign}\left(\sum_{i=1}^n \alpha_i \cdot y_i \cdot K(x_i, x_j) + b\right) \tag{9}$$

where $K(x_i, x_j)$ is the kernel function. The kernel trick is employed to transform the discriminator model from a linear into a non-linear case. Also, the Gaussian function is assumed as below:

$$K(x_i, x_j) = \exp\left(-\frac{1}{2\sigma^2}\|x_i - x_j\|^2\right) \tag{10}$$

where x_i, x_j stand for the input vectors, and σ symbolizes the kernel parameter [20].

Thus, in cases where multiple class problems arise, one class of the problem is usually compared with another class of the problem that usually generates the N classifier. In this regard, Vapnik [21] introduced a method that requires a solution obtained through the N quadratic programming optimisation problem. In this study, a beneficial method was referred to as 'one-versus-rest' method.

3.2.3 Hybrid MLP-SVM model

The combination of MLP and SVM network is predominantly aimed at linking the output layer of an MLP classifier by means of optimal margin hyperplanes. The idea of using a hybrid architecture of MLP-SVM is perceived to be a significantly improved model of MLPs in terms of performance [22]. That is, it discovers a distinctive solution to the last layer parameters by using convex optimisation along with a primal-dual understanding, as well as guaranteeing a higher bound in relation to testing errors. Moreover, the MLP-SVM model was applied with more efficiency in comparison to the non-linear SVMs which are trained in the input space. This was due to a non-linear SVM requiring selection and turning a kernel to attain a respectable non-linear mapping through the input space to a transformed feature space in which data were seemingly more linearly separable. Regarding the MLP-SVM method, this non-linear mapping was discretely optimised all through the MLP training in the system of the sigmoid kernel. In the stage of training in MLP, the function approximation application of this network was used to map the input and output data in the first layer. In order to optimise the network, back-propagation was used to minimise the relative entropy between the output delivery and the true label delivery by making use of the optimised input for hidden layer parameters. Then, the input data dimension was reduced to one-dimension space preparing the operation for its next stage. For the second training stage, the output data from MLP were taken and imported to SVM using the SVM scheme of "one-versus-rest" to predict the driver injury severity level [22, 23]. In Figure 3, the hybrid MLP-SVM model structure is displayed.

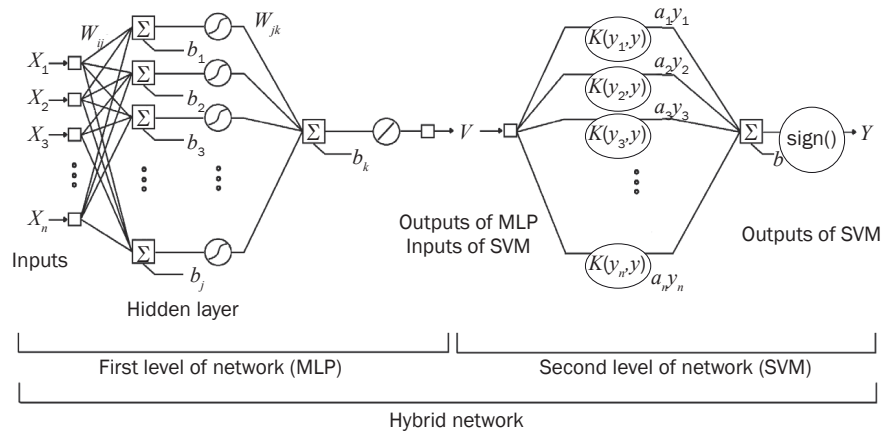


Figure 3 – A fully connected hybrid MLP-SVM structure

4. RESULTS AND DISCUSSION

4.1 Rank analysis using MLP and SVM prediction models

In this section, MLP and SVM models were applied to predict the injury severity by using MATLAB software as a powerful programming environment for each situation. Before processing the data into the prediction models that will then be used for comparison purposes, the factors that have slight influence on the prediction of the injury severity must be removed. This removal is to improve the quality of the data and lead to reduced input of the data dimension space that can help boost the speed of the networks, and consequently, this will affect the accuracy of the final predictions. Therefore, to reach this goal, the data were applied individually as an input for each model to compare the effect of each injury-related variable on the severity outcomes. Thus, the models predicted for 28 epochs by changing one input in each epoch while keeping other variables constant. It should be noted that the data were at the start shuffled as well as normalised for each repetition with the aim of having an equivalent series of values.

4.1.1 Sensitive predictors resulting from MLP

To pinpoint the more sensitive factors as well as to determine the rank for each input variable, the amount of error for each index resulting from MLP model was calculated. In fact, the sensitivity of the network against the absence of each input index for the presence of all indicators was reviewed.

Figure 4 shows the amount of Mean Square Error (MSE) versus the absence of each factor. Based on the obtained results of the MLP prediction model, the most significant contributory factors have been identified and ranked due to their MSE values. The range for MSE is specified nearly 15 – 21. Accordingly, based on the professional judgment, the boundary value of MSE was measured as 17% which is approximately 70% of the range. Therefore, the explanatory variables over

the threshold of 17% had a higher effect in the performance of the prediction. This means that these sensitive predictors were the most important accident-related variables that affected injury severity. In contrast, the application of the less reliable factors under 17% had less effect on the network performance than input data as well as indicating poor fitting of the model. Consequently, poor relationship between the injury severity and the crash-related factors were poor. In line with the significant results of the sensitivity test, the influence of the vehicle category was overhead compared to other indices listed in Figure 4 and played a huge role in the injury severity. This finding fits other studies wherein the type of vehicle was identified as the key role in the occurrence of injury severity [13, 14, 19]. The next main contributing factor refers to vehicle manoeuvre, which is attributed to actions immediately before the collision. This finding is in line with the previous studies showing that the increase of accidents is correlated with vehicle activities [3, 19].

A greater part of the accidents that took place were in the areas in which the crossing facilities were not available for the cyclists. Chang and Wang [19] found that the unprotected road users were at a greater risk of contributory injuries in collisions. The first point of impact is another factor that contributed to the accidents. This refers to the first point of contact made with another vehicle. A similar result was found by Abdel-Aty and Abdelwahab [14] who identified the point of impact as a sensitive predictor. Moreover, the factors attributed to the junction actions listed in the Figure 4 accounted as the other key influence predictors. In addition, the 2nd Road class was applied for junction accidents only; therefore, it appears that an alarming number of collisions in and around junctions was caused by drivers or riders. This predicament is an injury severity outcome which is consistent with the latest study by Curiel et al. [24], which discovered that almost half of the traffic crashes happened in five percent of the capital intersections. All the other findings are ranked in Figure 4.

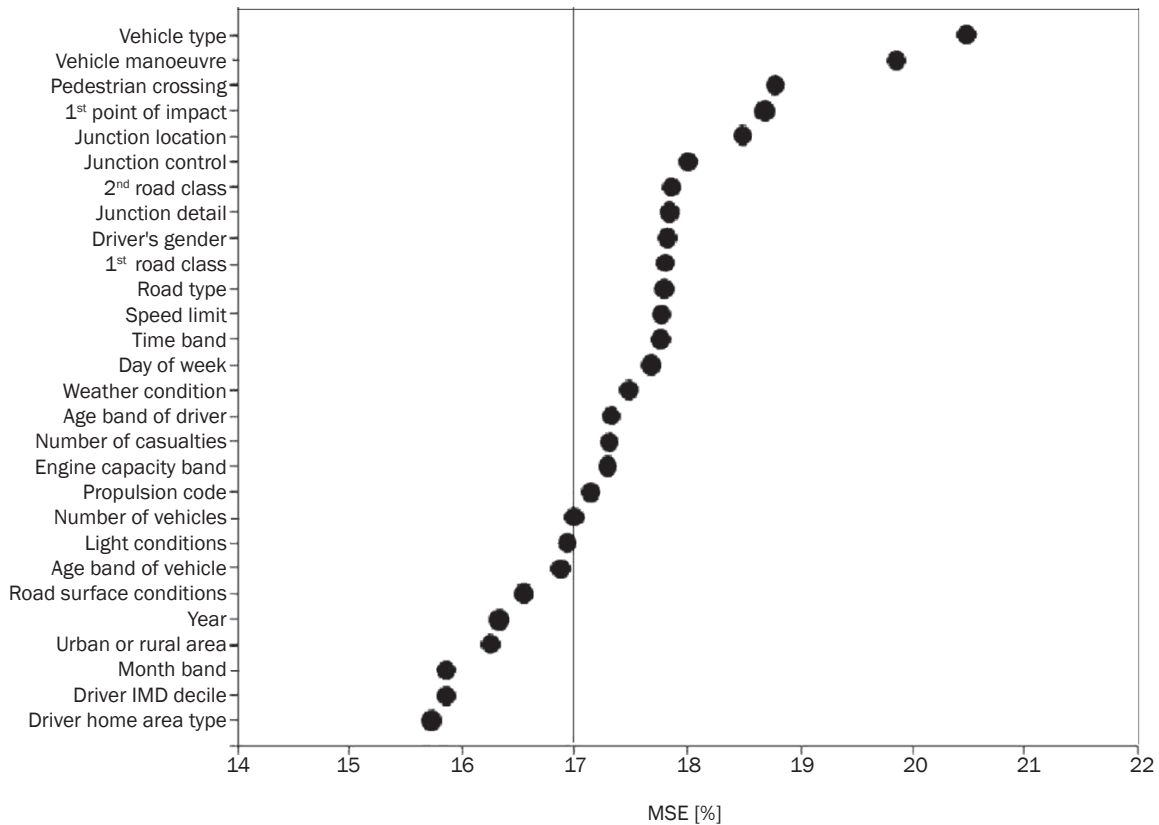


Figure 4 – Sensitivity of MLP

4.1.2 Sensitive predictors resulting from SVM

The SVM model was typically fit in the training and test sets. Moreover, the model was fit on a validation set in addition to the training and test sets. The amount of 70% of data was shared as the training data at random; 15% was taken for validation; and finally, the outstanding 15% was divided in the test data which are independent from the training data, but follow a similar likelihood of distribution. Correlation coefficient (R) value was used to assess the relation between the actual and the predicted levels [25]. Accordingly, all the R values have been evaluated for comparison between training, validation and testing data. Table 2 shows the values of R for each position. The closer the R is to 1 signifies greater correlation with the injury severity classes.

According to the sensitivity of SVM examination, the significant predictors have been detected. Likewise, the influences have been ranked by relying on the results in Table 2. The explanatory variables with greater R -value of 5.1% had more effect on the prediction performance. The key findings display that the most sensitive predictors are attributed to the same factors which were raised in the predictions resulting from MLP. The ranking of the major contributory factors related to vehicle type, manoeuvre, crossing facilities, point of impact, and junction actions are the same with MLP

outcomes. However, their impacts were nearly about triple in comparison to other contributing factors; all the additional variables almost being held near equal.

4.2 Requirement of road safety interventions

As result of the predictions by MLP and SVM, in order to generalise the dimensional feature space, the less sensitive predictors which had a minor effect in influencing the injury severity were dropped. Consequently, by eliminating the less significant parameters, the dimension space of input data was reduced to 19 labels for the implementation of final predictions.

The data were applied in the MLP and SVM models to detect sites and clusters that were required in the intervention of road safety. The outcomes from the above predictions along with the data analysis for the most important contributory factors which are shown in Table 3 display that the injuries resulting from cars and bicycles played a large part in the injuries. Following this, the most common injuries suffered by cyclists happened where the crossing facilities were not presented within 50 metres for the controlling of pedal riders crossing. Drivers and riders frequently contributed to the accidents through manoeuvres resulting from a poor turn or bend ahead. In addition, the impact of head-on accidents had severe consequences Furthermore, T/staggered junctions were definitely the

Table 2 – Sensitivity of SVM

Rank	Index name	Correlation coefficient (<i>R</i>) [%]			
		Training data	Validation data	Test data	All data
1	Vehicle type	89.4	86.8	85.1	88.6
2	Vehicle manoeuvre	87.6	73.6	72.6	76.9
3	Pedestrian crossing	85.6	81.3	83.7	85.2
4	1 st point of impact	82.1	78.9	78.6	81.3
5	Junction detail	79.7	75.7	74.4	78.9
6	Junction location	73.5	71.0	69.1	72.6
7	2 nd road class	71.1	69.6	58.8	70.3
8	Junction control	69.1	66.4	66.2	68.4
9	Driver's gender	26.5	25.4	24.9	26.1
10	Road type	23.3	21.5	21.0	22.8
11	Time range	22.9	21.9	21.3	22.7
12	Weather condition	18.4	17.9	17.4	18.2
13	1 st road class	11.4	10.4	10.1	11.2
14	Day of week	10.9	10.1	10.1	10.8
15	Speed limit	9.7	8.7	8.6	9.6
16	Driver or rider age	8.8	7.7	7.1	8.6
17	Engine capacity range	8.5	7.7	7.9	8.2
18	Propulsion	6.3	6.7	6.9	6.0
19	Number of casualties	5.4	4.9	4.1	5.1
20	Number of vehicles	3.9	2.8	2.6	3.2
21	Road surface condition	3.3	2.4	2.7	2.9
22	Light condition	3.7	2.5	2.5	2.9
23	Age range of vehicle	3.3	2.4	2.1	2.5
24	Urban or rural area	2.9	1.8	1.5	2.1
25	Driver home type area	2.8	1.8	1.7	2.2
26	Month band	2.5	1.4	1.3	1.9
27	Driver IMD decile	2.1	1.4	1.4	1.7
28	Year	1.2	0.5	0.5	0.8

capital's accident hotspots, particularly when they attempted to approach or were located in the mid of the junction leading into a C class road.

4.3 Second stage of the predictions using sensitive predictors

Following the examining of the models' sensitivity against the absence of each variable on the injury outcomes, in this Section, the models were predicted with nineteen more sensitive predictors using MLP, SVM, and hybrid MLP-SVM. MATLAB programming language was again used in the training, testing, and structural algorithms of these models. All the factors were normalised between zero and one. The delivered dataset was randomly separated to form two subsets: training data of 70% and test data of 30% which lead to the implementation of the structure optimisation

algorithm accompanied by the performance comparison between the models. Following this, one run was made using the random division. In the MLP network, nineteen neurons were considered for the input layer. Furthermore, in the hidden layer, the number of neurons was chosen by trial and error. In this layer, the tangent sigmoid function was used to find the relation between the input and output indices. For the process of training, the Levenberg-Marquardt backpropagation method was applied. Furthermore, the MSE was used as a performance point of the network and the training process stopped when the amount of MSE reached 0.005. As the structure of SVM is completely different from MLP, this learning machine was adjusted to pass the best hyperplane from different classes of data which has maximum margin between classes. Therefore, for the training of SVM, the Gaussian

Table 3 – Descriptive statistics of the most sensitive predictors

Variable	Label	Total [%]
Vehicle type	Car	36.66
	Pedal cycle	26.19
	Motorcycle	15.60
	Public service vehicles	08.20
	Light goods vehicles	07.65
	Heavy goods vehicles	05.25
	Other vehicles	00.46
Vehicle manoeuvre	Waiting or turning right at a junction	47.09
	Driving or riding through bend	19.56
	Overtaking	10.97
	Held up or waiting to drive	06.14
	Moving off	05.03
	Stopping or slowing down	04.31
	Parked	02.57
	Moving line of traffic	02.46
	Reversing	01.88
Crossing services	Phase at signal intersection	46.70
	No crossing services across fifty metres	44.75
	Light organised crossing	04.92
	Zebra	02.37
	Pedestrian island	00.69
	Overpass or underpass	00.57
1 st point of impact	Front	39.86
	Nearside	23.74
	Offside	19.17
	Back	11.37
	Did not impact	05.86
Junction detail	3-way or staggered intersection	54.78
	Cross road	20.22
	Not at intersection or across twenty metres	15.20
	Over four roads meet	04.48
	Roundabout	02.94
	Other intersection	02.38
Junction location	Approaching or holding up	49.29
	Middle of intersection	31.30
	Not at or across twenty metres of intersection	15.25
	Entering or leaving major road	03.97
	Entering or leaving roundabout	00.17
	Entering from slip road	00.03
2 nd road class	C	53.42
	A	29.40
	Unclassified	14.79
	B	02.39
Junction control	Smart traffic light	47.08
	Uncontrolled or give way	36.09
	Not at intersection or across twenty metres	15.03
Injury severity classes	C4 No injury (property damage only)	57.20
	C3 Slight injury	37.49
	C2 Severe injury	05.14
	C1 Fatal injury	00.17

Table 4 – ACC, Error, SEN, and confusion matrix for each model

Model	Training				Test			
	ACC [%]	Error [%]	Confusion matrix	SEN [%]	ACC [%]	Error [%]	Confusion matrix	SEN [%]
MLP	84.41	15.59	$\begin{bmatrix} 1 & 3 & 0 & 0 \\ 1 & 6 & 119 & 0 \\ 0 & 18 & 767 & 134 \\ 0 & 0 & 107 & 1295 \end{bmatrix}$	C1 25.00 C2 04.76 C3 83.46 C4 92.37	83.82	16.18	$\begin{bmatrix} 0 & 2 & 0 & 0 \\ 0 & 2 & 52 & 0 \\ 0 & 10 & 315 & 69 \\ 0 & 0 & 37 & 546 \end{bmatrix}$	C1 00.00 C2 03.70 C3 79.95 C4 93.84
SVM	81.68	18.32	$\begin{bmatrix} 0 & 1 & 3 & 0 \\ 0 & 0 & 110 & 16 \\ 0 & 0 & 738 & 181 \\ 0 & 0 & 156 & 1246 \end{bmatrix}$	C1 00.00 C2 00.00 C3 80.30 C4 88.87	80.88	19.12	$\begin{bmatrix} 0 & 0 & 2 & 0 \\ 0 & 0 & 53 & 1 \\ 0 & 0 & 325 & 69 \\ 0 & 0 & 76 & 525 \end{bmatrix}$	C1 00.00 C2 00.00 C3 82.49 C4 87.35
Hybrid MLP-SVM	91.35	08.65	$\begin{bmatrix} 3 & 1 & 0 & 0 \\ 2 & 14 & 110 & 0 \\ 2 & 52 & 827 & 28 \\ 0 & 0 & 7 & 1395 \end{bmatrix}$	C1 75.00 C2 11.11 C3 89.99 C4 99.50	90.58	09.42	$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 4 & 6 & 44 & 0 \\ 0 & 30 & 348 & 16 \\ 0 & 0 & 4 & 597 \end{bmatrix}$	C1 50.00 C2 11.11 C3 88.32 C4 99.33

function had been selected to be an effective kernel trick function. For the hybrid MLP-SVM, at the first level, the MLP had reduced the dimension space of input data from nineteen to one, and it led to facilitate the process of prediction for the second layer (SVM). The obtained results from each model are shown in Table 4 using the confusion matrix. Also, for the evaluation of each model, in addition to accuracy (ACC), error parameters, and sensitivity (SEN) were used in order to display the rate of true positive prediction of each class (C), where the classes consisted of C1 (fatal injury), C2 (serious injury), C3 (slight injury), and C4 (damage only).

4.4 Classes resulting from the prediction models

According to the outcomes from the confusion matrix shown in Table 4, the comparison of the actual and the predicted classes of the injury severity training and test data was applied and are explained as follows:

In the MLP model, the accuracy measure for the prediction in the training and test phase were found to be 84.41% and 83.82%, respectively. However, as the number of the fatalities was not particularly high and there were only six pedal riders, due to the lack of data related to this class, the network was able to evaluate only one accurate prediction for C1 class in the training process. Therefore, in the test phase, the amount of sensitivity for this class equalled zero. For C2, the network was unable to perform the prediction of the driver injury severity. In fact, in the training process, the network accommodated the input parameters with C3 instead of C2. For class C3, the classification was almost desirable and the amount of sensitivity for the training and test phase was obtained around 80%. Furthermore, in this class, the classification tended more to C4 in lieu of C2. Finally, among the classes,

the prediction of the injury severity for C4 was performed better than for other classes (with more than 90% sensitivity).

In the SVM model, the amount of accuracy obtained was slightly less than the MLP network (81.63% for the training phase and 80.88% for the test phase). Nonetheless, for C1, both in the training and the test process, the SVM had very poor performance (sensitivity equalled zero). As for the prediction of C2, the SVM was unsuccessful and more data were incorrectly classified. As for C3 and C4, compared to the two previous classes, the performance of SVM had improved and the sensitivity value for the training and test process was obtained around 80% and 88%, respectively.

The hybrid model provided a superior fitting of the model compared to the MLP and SVM models and the obtained results have been very satisfactory. However, the lack of data for C1 and incorrect classification for C2 still remain. Finally, the hybrid model was able to maximise the accuracy rate for predicting the injury severity by 10%.

A similar comparison result was achieved by Bellili et al. (2003) in a classification task. The goal of their study was to reduce the recognition of error rate using the hybrid MLP-SVM recogniser. Accordingly, their outcome proved that the hybrid MLP-SVM model significantly increased the performance [22].

5. CONCLUSION

This study mainly concentrated on the potential improvements in the driver or rider injury severity prediction by using MLP, SVM, and a combination of these two models. According to the statistics which is connected to road casualties of all severities in the city of London, these collision reduction techniques estimated the maximum probability of the injury severity classes. Moreover, the result of the prediction models led to improved comprehension of the significant

relation among severity of injuries sustained by drivers or riders and the related explanatory variables that contribute to the collisions. Several sensitive predictors were identified which are believed to have contributed to the injury severities.

The main contributory factor refers to the vehicle type. Accordingly, cars and bicycles presented the highest concentration of injuries. Road users often contributed to the collisions through manoeuvre as a result of poor turn or bend. The next major factor was found to be the insufficient number of pedestrian crossing facilities for cyclists. Following this, the first point of impact was another factor. Thus, the severity of a head-on crash was the most alarming type of accident in which the road users were involved. Further to this, the indices associated with the junction actions were other contributory factors in which T/Staggered junctions were identified as the worst performing intersections.

This research ends by performing the second stage of the predictions with the objective of maximising the model accuracies. As a result, a comparison between MLP, SVM, and the proposed hybrid method was employed by using the most important contributing factors. Following this, the differences between the MLP, SVM and Hybrid MLP-SVM were outlined, and their performances in relation to accuracy of the predictions were compared. The findings of this comparison concluded that the MLP model offered a similar overall accuracy of the SVM model, but the Hybrid model achieved a superior injury severity prediction with higher accuracy. Likewise, the testing classification accuracies were 83.82% for MLP, 80.88% for SVM, and 90.58% for the Hybrid model.

The analysis obtained from the confusion matrices showed that the lack of data for C1 led to poor classification. Furthermore, the best class was C4 due to the number of data for this class being sufficient in the training phase. For C2, quantitative effects of each input variable on the injury severity could not be predicted well and it tended to work with C3. Therefore, in terms of future work, in order to achieve better results, C2 and C3 should be merged together.

The outcome of this research can be used as a function to evaluate the safety performance of the overall transport networks. Having this methodology as a tool, the engineers can point out the safety problems within the system and try to minimise them by efficiently allocating the resources available for the decision makers. Subsequently, the results of this study could lead to the implementation of an Advanced Automatic Collision Notification (AACN) technology. This system works by recognising the key influencing variables in real-time in conjunction with the analysis of available crash data. Using an AACN system within the traffic network, a warning could be announced in advance in order to reduce the risk of injury.

ACKNOWLEDGEMENTS

The first author wishes to thank the DFT Road Safety Statistics Team for providing the data used in this research.

دکتر میثم صیامی دودران (نویسنده مسئول)
گروه مهندسی عمران دانشگاه مدیترانه شرقی،
قبرس

دانشکده اقتصاد، مدیریت و فناوری اطلاعات
دانشگاه فنی دولتی وارونژ، فدراسیون روسیه
در عین حال او عضو رسمی تیم ایمنی راه در
دولت بریتانیای کبیر می باشد

دکتر ارسون ایشچی اوغلو
گروه کامپیوتر و فناوری آموزشی دانشگاه
مدیترانه شرقی قبرس

**پیش بینی شدت آسیب تصادفات ترافیکی با
استفاده از یک سری شبکه های عصبی
مصنوعی: یک مطالعه موردی برای شهر لندن**

چکیده: در این مقاله، پیش بینی شدت آسیب راننده، موتور یا دوچرخه سوار در تصادفات ترافیکی در مرکز شهر لندن مورد بررسی قرار گرفته است. بدین منظور، مدل های پرسپترون چند لایه، ماشین بردار پشتیبان و یک روش نوین که ترکیبی از دو مدل ذکر شده می باشد مورد استفاده قرار گرفته است. با کشف همبستگی بین کلاس های شدت آسیب و عواملی که در ایجاد آنها مؤثر بوده، این تکنیک پیشگیری قادر به شناخت عوامل مؤثر بر روی شدت آسیب می باشد. نتایج بدست آمده حاکی از آن است که نوع وسیله نقلیه، مانور وسیله نقلیه، عدم وجود امکانات لازم برای عبور دوچرخه سواران، برخورد اولیه در لحظه تصادف و واکنش های انجام شده در چهارراه ها دارای بیشترین تأثیرات بر احتمال شدت آسیب دیدگی می باشند. سپس با هدف به حداکثر رساندن صحت پیش بینی، مقایسه ای بین مدل ها انجام گردید. در این مرحله با بکارگیری حساس ترین پیش بینی کننده ها، عملکرد مدل ها در برابر یکدیگر ارزیابی گردید. نتایج بدست آمده حاکی از آن می باشد که کارآرایی مدل ترکیبی به میزان قابل توجهی نسبت به مدل های دیگر افزایش یافته و مدل پیشنهادی پیشرفتی در حدود نود درصد را در دقت پیش بینی کسب نموده است.

کلمات کلیدی: ایمنی جاده؛ تصادفات ترافیکی؛ پیش بینی شدت آسیب؛ عوامل مؤثر

REFERENCES

- [1] Xie Y, Lord D, Zhang Y. Predicting Motor Vehicle Collisions using Bayesian Neural Networks: An Empirical Analysis. *Accident Analysis & Prevention*. 2007;39(5): 922-933.
- [2] DFT (Department for Transport). *Instructions for the Completion of Road Accident Reports from non-CRASH Sources*; 2011. Available from: <https://www.gov.uk/government/uploads/system/uploads/>

- attachment_data/file/230596/stats20-2011.pdf [Accessed 28th June 2016].
- [3] Li Y, Ma D, Zhu M, Zeng Z, Wang Y. Identification of significant factors in fatal-injury highway crashes using genetic algorithm and neural network. *Accident Analysis & Prevention*. 2018;111: 354-63.
- [4] Tabachnick BG, Fidell LS. *Using multivariate statistics*. 6th Edition. Boston, MA: Pearson; 2012.
- [5] Karlaftis MG, Vlahogianni EI. Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. *Transportation Research Part C: Emerging Technologies*. 2011;19(3): 387-99.
- [6] Principe JC, Euliano NR, Lefebvre WC. *Neural and adaptive systems: fundamentals through simulations*. New York: Wiley; 2000.
- [7] Abdelwahab HT, Abdel-Aty MA. Artificial neural networks and logit models for traffic safety analysis of toll plazas. *Transportation Research Record*. 2002;1784(1): 115-25.
- [8] Li X, Lord D, Zhang Y, Xie Y. Predicting motor vehicle crashes using support vector machine models. *Accident Analysis & Prevention*. 2008;40(4): 1611-8.
- [9] Zhang HH. Variable selection for support vector machines via smoothing spline ANOVA. *Statistica Sinica*. 2006;16(2): 659-674.
- [10] Li Z, Liu P, Wang W, Xu C. Using support vector machine models for crash injury severity analysis. *Accident Analysis & Prevention*. 2012;45: 478-486.
- [11] Zhang J, Li Z, Pu Z, Xu C. Comparing Prediction Performance for Crash Injury Severity Among Various Machine Learning and Statistical Methods. *IEEE Access*. 2018;6: 60079-60087.
- [12] Zeng Q, Huang H. A stable and optimized neural network model for crash injury severity prediction. *Accident Analysis & Prevention*. 2014;73: 351-358.
- [13] Abdelwahab HT, Abdel-Aty MA. Development of artificial neural network models to predict driver injury severity in traffic accidents at signalized intersections. *Transportation Research Record*. 2001;1746(1): 6-13.
- [14] Abdel-Aty MA, Abdelwahab HT. Predicting injury severity levels in traffic crashes: a modeling comparison. *Journal of Transportation Engineering*. 2004;130(2): 204-210.
- [15] Delen D, Sharda R, Bessonov M. Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. *Accident Analysis & Prevention*. 2006;38(3): 434-44.
- [16] Yu R, Abdel-Aty M. Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data. *Safety science*. 2014;63: 50-56.
- [17] Chen C, Zhang G, Qian Z, Tarefder RA, Tian Z. Investigating driver injury severity patterns in rollover crashes using support vector machine models. *Accident Analysis & Prevention*. 2016;90: 128-139.
- [18] Iranitalab A, Khattak A. Comparison of four statistical and machine learning methods for crash severity prediction. *Accident Analysis & Prevention*. 2017;108: 27-36.
- [19] Chang LY, Wang HW. Analysis of traffic injury severity: An application of non-parametric classification tree techniques. *Accident Analysis & Prevention*. 2006;38(5): 1019-1027.
- [20] Burges CJ. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*. 1998;2(2): 121-167.
- [21] Vapnik VN. *The nature of statistical learning theory*. Springer; 1995.
- [22] Bellili A, Gilloux M, Gallinari P. An MLP-SVM combination architecture for offline handwritten digit recognition. *Document Analysis and Recognition*. 2003;5(4): 244-252.
- [23] Bishop CM. *Neural networks for pattern recognition*. Oxford University Press; 1995.
- [24] Curiel RP, Ramírez HG, Bishop SR. A novel rare event approach to measure the randomness and concentration of road accidents. *PLoS ONE*. 2018;13(8): e0201890.
- [25] Siamidoudaran M, Iscioglu E, Siamidodaran M. Traffic injury severity prediction along with identification of contributory factors using learning vector quantization: A case study of the city of London. *SN Applied Sciences*. 2019 Oct;1(10): 1268. Available from: doi:10.1007/s42452-019-1314-6