Journal of Civil Engineering and Urbanism

Volume 5, Issue 5: 193-199; September 25, 2015 ISSN-2252-0430

Application of Support Vector Machine for Crash Injury Severity Prediction: A Model Comparison Approach

Iman Aghayan[1], Mansour Hadji Hosseinlou[2], Mehmet Metin Kunt[3]

[1]*Department of Civil Engineering, Shahrood University of Technology, Shahrood, Iran*
[2]*Department of Civil Engineering, K.N. Toosi University of Technology, Tehran, Iran*
[3]*Department of Civil Engineering, Eastern Mediterranean University, Gazimagusa, Mersin 10, Turkey*
*Corresponding authors' E-mail: iman.aghayan@shahroodut.ac.ir; metin.kunt@emu.edu.tr

ABSTRACT: The study presented in this paper investigated the application of using support vector machine with different kernel functions for crash injury severity prediction. A support vector machine model was developed for predicting the injury severity related to individual crashes based on crash data. The models were developed using the input parameters of driver's age and gender, the use of a seat belt, the type and safety of a vehicle, weather conditions, road surface, speed ratio, crash time, crash type, collision type and traffic flow. Also, three injury levels were considered as output parameters for this study (i.e. no injury, evident injury and fatality). The overall prediction accuracy of the support vector machine model was compared to the multi-layer perceptron, genetic algorithm, combined genetic algorithm and pattern search. The results demonstrated that the constructed multi-layer perceptron's performance was slightly better than the support vector machine for injury severity prediction. Whereas, support vector machine involves much lower computational cost than multi-layer perceptron because of using a straight forward algorithm in learning phase. The percent of prediction accuracy for the multi-layer perceptron model was 86.2%, which was higher than the support vector machine model with polynomial kernel (81.4%) followed by the combination of the genetic algorithm and pattern search (78.6%) and genetic algorithm (78.1%). The classification results of the two-level (no-injury and evidence injury/fatality) support vector machine found to be 85.3% was higher than multi-class classification (81.4%).

Keywords: Crash Injury Severity Prediction, Genetic Algorithm, Multi-Layer Perceptron, Pattern Search, Support Vector Machine

ORIGINAL ARTICLE
PII: S225204301500031-5
Received 13 Aug. 2014
Accepted 25 Jan. 2015
Revised 17 June 2015

INTRODUCTION

The number of traffic crashes worldwide is increasing. Many efforts have been conducted to reduce the crash occurrence. One of the most important tools for investigating the relationship between crash occurrence and traffic risk factors is a crash prediction model. Traditional measures to reduce crashes include improved geometric design, congestion management strategies and better driver education and enforcement. While these measures are generally effective, they are often not feasible or prohibitively expensive to implement. Road traffic crashes are usually caused by the composite actions of humans, vehicles, road, and weather, and their outcomes often involve casualties and economic loss. The relationship between a crash and the influencing factors is nonlinear and complicated; it cannot be described with an explicit mathematical model. Among them, the negative binominal (NB) model arises mathematically (and conveniently) by assuming that unobserved crash heterogeneity (variation) across sites (intersections, road segments, etc.) is Gamma distributed while crashes within sites are Poisson distributed (Washington et al., 2003). Bayesian empirical methods have also been developed (Ng and Sayed, 2004; Wright et al., 1988). Poisson, Poisson-Gamma (NB) and other related models are called generalized linear models. Hadji Hosseinlou and Aghayan (2009) used fuzzy logic to predict traffic crash severity on the Tehran-Ghom freeway in Iran. Aghayan et al. (2013) investigated Fuzzy c-means (FCM) clustering based on clustering algorithms for traffic crash in Cyprus.

Support Vector Machines (SVMs) have been introduced as a new and novel machine learning technique according to the statistical learning theory. SVMs developed by Vapnik (1995) are used for classification and regression problems. Structural Risk Minimization (SRM) applied by SVM can be superior to Empirical Risk Minimization (ERM) since SRM minimizes the generalization error. SVMs have scarcely been used as a modelling approach in the analysis of crash-related injury severity. Lv et al. (2009) used SVMs for real-time highway crash prediction. They tried to find out traffic conditions leading to traffic crashes more likely using the SVM method by considering the geometry, environmental factors, etc. Data was collected from the simulation software TSIS. According to the results obtained from SVM model, hazardous traffic conditions cannot be identified from normal traffic conditions with regard to one single variable. Li et al. (2008) used SVM models for predicting motor vehicle crashes. NB regression and SVM models were developped and compared using data collected on rural frontage roads in Texas. The results showed that SVM models predict crash data more effectively and accurately than traditional NB models.

ANNs have been verified to be efficient in many fields. Neural networks are commonly used for non-

To cite this paper: Aghayan I., Hadji Hosseinlou M., Metin Kunt M. 2015. Application of Support Vector Machine for Crash Injury Severity Prediction: A Model Comparison Approach. *J. Civil Eng. Urban.*, 5 (5): 193-199.
Journal homepage http://www.ojceu.ir/main/

linear modelling and forecasting. In traffic safety, some studies have applied ANNs to predicting crash rates and analysing crashes, but none have used twelve parameters, including important factors with detail. Thus, this study attempted to incorporate all relevant parameters into the models to achieve a high percentage of crash forecasting. Mussone et al. (1999) applied ANNs to analyse vehicular crashes that occurred at an intersection in Milan, Italy. A number of studies have attempted to identify groups of drivers at a greater risk of being injured or killed in traffic crashes (Zhang et al., 2000; Valent et al., 2002).

Bedard et al. (2002) applied multivariate logistic regression analysis to investigate the effects of a driver, crash and vehicle characteristics on fatal crashes. Lord et al. (2005) conducted analysis on the relationship among crash, density (vehicles per km per lane) and v/c ratio. They found that along with an increase in v/c ratio, fatal and single-vehicle crashes decreased after some point, and crash rates followed U-shaped relationship. More recent applications in the transportation field using the ANN have included traffic prediction (Yin et al., 2002; Zhong et al., 2004), the estimation of traffic parameters (Tong and Hung, 2002), traffic signal control (Zhang et al., 2001), incident detection (Jin et al., 2002; Yuan and Cheu, 2003), travel behaviour analysis (Subba Rao et al., 1998; Hensher and Ton, 2000; Vythoulkas and Koutsopoulos, 2003) and traffic crash analysis (Mussone et al., 1996; Mussone et al., 1999; Sohn and Lee, 2003; Abdel-Aty and Pande, 2005). For example, Abdelwahab and Abdel-Aty (2001) used ANNs for modelling the relationship between driver injury severity and crash factors related to the driver, vehicle, roadway, and environmental characteristics. Their study focused on classifying crashes into one of three injury severity levels using the readily available crash factors. These authors limit their domain of study to two vehicle crashes that occurred at intersections with signals. The predictive performance of a Multi-Layer Perceptron (MLP) neural network was compared to the performance of the ordered logit model.

The obtained results showed that MLP achieved better classification (correctly classifying 65.6 and 60.4% of cases for training and testing phases respectively) than the ordered logit model (correctly classifying 58.9 and 57.1% of cases for training and testing phases respectively). Aghayan et al. (2012) applied FCM and Fuzzy Subtractive (FS) clustering compared with ANN by considering accuracy and response time criteria. The results represented that ANN can be the appropriate model for prediction accuracy and the lowest response time was achieved by FS algorithm in comparison with the applied models.Genetic Algorithms (GAs) are powerful stochastic search techniques based on the principle of natural evolution. These algorithms were first introduced and investigated by John Holland (1975). According to Chang and Chen (2000), regression models generated by genetic programming (GP) are also independent of any model structure.

According to Deschaine and Francone (2004), the GP is observed to perform better than classification trees with lower error rates and also outperforms neural networks in regression analysis. Several studies (Park et al., 2000; Ceylan and Bell, 2004; Teklu et al., 2007) have used GP methods in the traffic signal system and network optimization. Kunt et al. (2011) used ANN, GA and GA combined with Pattern Search (PS) for predicting the severity of freeway traffic crashes. The performance of these methods was compared to find the most suitable method for predicting crash severity. The results showed that the ANN provided the best prediction.The main aim of this research is to investigate the application of SVM for crash injury severity prediction. MLP, GA, and combined GA and PS models are compared with SVM model to effectively evaluate the classification capability of SVMs. In addition, the most accurate one is selected according to twelve input parameters and three levels of injury severity.

Data description

The dataset used in this study was derived from a total of 1063 reported traffic crashes in Tehran, the capital of Iran. These crashes were selected from the total number of crashes that occurred on the Tehran-Ghom freeway in 2007 since these were the only complete crash records. These data were used as training and testing data for the SVM, MLP, GA and combined GA and PS methods.

Three injury levels were considered for this study (i.e. no injury, evident injury and disabling injury/fatality), and twelve variables were selected from the obtained data. The vehicle speed in police reports was calculated by a camera or breaking distance. Speed ratio was used as one of the input variables defined as the ratio of estimated speed at the time of a crash to posted speed limit at the crash location. Road geometry parameters were not taken into consideration because the selected road had a desirable geometry common to all crashes in the dataset.

Because the data have dissimilar units and magnitudes, the data for each variable had to be normalized. Data normalization can improve the data fitting as well as prediction performances and is required for input into the models. Table 1 shows input and output variables.

MATERIAL AND METHODS

The SVM model treats the traffic crash modelling as a classification problem. The SVM can be used to determine the suitability of those input variables and injury severity levels for model predictions. For comparison purposes, GA, combined GA and PS, and a MLP neural network were developed according to the same dataset (Kunt et al., 2011).

The performance of these methods was compared to find the most suitable method for predicting crash severity at three levels: fatality, evident injury, and no injury.

Support vector machine model

SVMs have been introduced as a new and novel machine learning technique according to the statistical learning theory. The basic SVM considers two-class pattern recognition problems. The basic idea of SVM is to find a separating hyperplane between the classes in the N-dimension space of the inputs. The largest margin between points of the different classes can result in the better generalization error of classifier.

Table 1. A description of study variables

Parameters	Subdivided Parameters	Variable	Coding/Values	Data
Input Variables				
1	2	Driver's Gender	Man= (1,0)	97.56%
			Woman= (0,1)	2.44%
2	1	Driver's Age	Year	20-34=39%
				35-49=44%
				50-64=10%
				65-79=7%
3	2	Use of Seat Belt	In use= (1,0)	78.66%
			Not in use= (0,1)	21.34%
4	3	Type of Vehicle	Passenger car= (1,0,0)	83.54%
			Bus= (0,1,0)	2.44%
			Pick-up= (0,0,1)	14.02%
5	2	Safety of Vehicle	High standard= (1,0)	31.71%
			Low standard= (0,1)	68.29%
6	4	Weather Condition	Clear= (1,0,0,0)	56.71%
			Snowy= (0,1,0,0)	7.93%
			Rainy= (0,0,1,0)	10.37%
			Cloudy=(0,0,0,1)	25%
7	3	Road Surface	Dry= (1,0,0)	75%
			Wet= (0,1,0)	17.68%
			Snowy/Icy= (0,0,1)	7.32%
8	1	Speed Ratio	km/hr / km/hr	
9	2	Crash Time	Day= (1,0)	65.85%
			Night= (0,1)	34.15%
10	2	Crash Type	With vehicle= (1,0)	74.81%
			With multiple vehicles= (0,1)	25.19%
11	3	Collision Type	Rear-end= (1,0,0)	51.95%
			Right-angle= (0,1,0)	30.24%
			Sideswipe= (0,0,1)	17.80%
12	1	Traffic Flow	veh/h	
Output variables				
1	3	Driver Injury Severity	Fatality= (1,0,0)	14.02%
			Evident injury= (0,1,0)	38.41%
			No injury= (0,0,1)	47.56%

The basic SVM formulation solves the binary classification problems; thus, several binary classifiers should be applied for constructing a multi-class classifier or making fundamental changes to the original formulation to consider all classes at the same time. The binary classifiers for both linear and nonlinear separable data are mentioned below.

Training samples are considered as:

$$X = \{(x, y)|(x_1, y_1), (x_2, y_2), \dots, (x_k, y_k)\}, \quad k=1,\dots, N \qquad (1)$$

where $x_k \in R^d$ is the kth input pattern, d denotes the dimension of the input space and y_k is its corresponding observed result, which is a binary variable 1 or -1. Here, x_k denotes attributes of organization and y_k is observed result of whether the injury severity is no injury or fatality/evident injury. Therefore, if the injury severity causes fatality/evident injury then $y_k = -1$, else $y_k = 1$. Considering that the training set is linearly separable after being mapped into higher dimensional feature space by nonlinear function $\emptyset(\bullet)$. Thus, the classifier can be constructed as:

$$w^T \emptyset(x_k)+b \geq 1 \qquad \text{if } y_k = 1$$
$$w^T \emptyset(x_k)+b \leq -1 \qquad \text{if } y_k = -1 \qquad (2)$$

The distance between the two boundary lines is $\frac{1}{2}\|w\|^2$. The maximal margin classifier optimizes this by separating the data with the maximal margin hyperplane. Meanwhile, the training set is usually not linearly separable even mapped into a high dimensional feature space; thus, a perfect separating hyperplane cannot happen to make each x_k satisfy condition (Eq. 2). Consequently, soft margin SVM is used to penalize misclassification errors and to employ a parameter (the soft margin C) to control the cost of misclassification. In the constraints, the positive slack variable $(\varepsilon_k)$is introduced to measure how much the margin constraints are violated (Vapnik, 1995).

$$\min_{w,b,\varepsilon_k} \frac{1}{2} w^T w + C \sum_{k=1}^{N} \varepsilon_k$$
$$\text{subject to } y_k[w^T \emptyset(x_k) + b] \geq 1 - \varepsilon_k, k = 1, \dots, N$$
$$\varepsilon_k \geq 0 \qquad (3)$$

where C is the regularizing (margin) parameter or penalty factor that determines the trade-off between the maximization of the margin and minimization of the classification error (Gun, 1998; Cristianini and Shawe-Taylor, 2000).

This constraint along with the objective of minimizing function can be solved using LaGrange multipliers (α_i). Thus, using LaGrange multipliers techniques can lead to the following dual optimization problem.

$$\min P(\alpha) = -\sum_{k=1}^{N} \alpha_k + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$
$$\text{subject to } \sum_{k=1}^{N} \alpha_k y_k = 0, \quad 0 \leq \alpha_k \leq C, k = 1, \dots, N \qquad (4)$$

$K(x_i, x_j) \equiv \emptyset(x_i)^T \emptyset(x_j)$ is called the kernel function. After solving Eq. 4 and substituting $w = \sum_{k=1}^{N} \alpha_k y_k \emptyset(x_k)$ into original classification problem, the following classifier is obtained:

$$y(x) = \mathrm{sign}(w^T \emptyset(x) + b)$$
$$= \mathrm{sign}\left(\sum_{k=1}^{N} \alpha_k y_k K(x, x_k) + b\right) \qquad (5)$$

There are different kernel functions that can be used for traffic crash analysis. In this study, linear function, polynomial function (5 degree), Radial Basis Function (RBF), and sigmoid function were applied so as to find the best kernel function, as shown in Table 2. The most important factors that influence the SVM's performance are the kernel parameters and the penalty factor. To achieve a better classification effect, the values of parameters in each model are important. These parameters are penalty factor C and kernel parameters (such as γ in RBF). The leave-one-out n-fold (5-fold) cross validation is a procedure to determine the best hyper parameters (C, γ, d, r) for SVM. For each crash record, the training dataset has a label which indicates the severity level (i.e. no injury, evident injury and fatality) and its paired individual crash data (input variables). Based on training crash data, the relationship between injury severity and input variables were learned in SVM model. For obtaining the maximum performance of model, the optimal values of the parameters were estimated. Based on the crash information defined in the testing crash dataset, the SVM model can make prediction on the severity level of each crash. The predicted and observed severities of crashes can be compared to evaluate the accuracy (the proportion of the total number of predictions that are correct) of correctly classified crashes.

Table 2. List of popular kernel functions

Type of classifier	Kernel function
Linear kernel	$K(x_i, x_j) = x_i^T x_j$
Polynomial kernel	$K(x_i, x_j) = (\gamma x_i^T x_j + r)^d$
Radial basis kernel	$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$
Sigmoid kernel	$K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$

$\gamma, r, and\ d$ are kernel parameters.

The SVM model demonstrated so far is for two-category classification. However, this model can be extended to multi-category classification tasks. One-versus- rest (OVR) approach is used to solve multi-level injury severity classification problem. In order to locate the best hyper parameters, leave-one-out cross validation is considered for SVM light (Joachims, 1998) and LIBSVM (Chang and Lin, 2013) in the MATLAB software. In this code, we want to illustrate how to perform classification using n-fold cross validation, which is a common methodology to use when the data set does not have explicit training and testing set separately. Such data sets usually come as a single set and they need to be separated into n equal parts/folds. The leave-one-out n-fold cross validation is to classify observations in a fold k by using the model trained from {all}-{k} models, and repeat the process for all k. The overall accuracy is obtained by averaging the accuracy per each of the n-fold cross validation. The observations are separated into n folds equally, the code use n-1 folds to train the SVM model which will be used to classify the remaining 1 fold according to standard OVR. The leave-one-out procedure should therefore be efficient for small sample sizes. Also for evaluation of the obtained results from classified data, the confusion matrix is used and is defined as an error matrix or a contingency table to determine the performance of the network. A grid searching algorithm was used to determine the kernel parameters related to the SVM model.

In this study, multi-class and two-class classification problems were considered. For achieving the better predictions, multi-class classification was reduced to a two-class classification. Thus, the three injury severities were converted to two severity levels (no-injury and evidence injury/fatality level).

Multi-Layer Perceptron Model

This study used a MLP neural network architecture that consisted of a multi-layer feed-forward network with sigmoid hidden neurons and linear output neurons as well as a network that was trained with the Levenberg-Marquardt back-propagation algorithm. The MLP model consisted of two layers, with each layer having a weight matrix W, a bias vector b, and an output vector p^i, with $i > 1$. Figure 1 shows the selected final prediction model for each layer in the MLP model where the number of the layer is appended as a superscript to the variable. For the different weights and other elements of the network, superscripts were applied to recognize the source (second index) and the destination (first index). Layer weight (LW) matrices and input weight (IW) matrices were used in the MLP model. The model was applied to the data that were randomly divided into sets for model training, testing, and validating. The MLP model had 12 inputs, 25 neurons in the first layer, and 3 neurons in the second layer. The output layer of the MLP model consisted of three neurons representing the three levels of injury severity. Of the original data, 70% were used in the training phase. While the validation and test data sets each contained 15% of the original data. A constant input 1 was fed to the bias for each neuron with regard to the outputs of each intermediate layer that were the inputs to the following layer. Thus, layer 2 could be analysed as a one-layer network with 25 inputs, 3 neurons, and a 3*25 weight matrix W^2; in such circumstances, the input layer 2 is p^2. All vectors and matrices of layer 2 have been identified; the layer can be treated as a single-layer network on its own. However, the objective of this network is to reduce the error e through the Least Mean Square (LMS) error algorithm that calculates the difference between t and p^i in which $i > 1$ and t is the target vector. The perceptron learning rule calculates the desired changes (target output) to the perceptron's weights and biases, given an input vector p^1 and the associated error e (Kunt et al., 2011).

RESULTS AND DISCUSSION

This study applied the SVM to predict the severity of traffic crashes. For comparison purposes, GA, combined GA and PS, and MLP neural network were developed (Kunt et al., 2011).

MATLAB software was used for comparing the performance of three modelling approaches (ANN, GA, and combined GA and PS) discussed earlier. And, the LIBSVM, SVM light and SVM multiclass were applied for the SVM model. The grid searching method was considered and the best values related to kernel parameters were selected automatically using the software.

The multi-class classification results of the SVM for different kernel functions by using LIBSVM are graphically depicted in Figure 2. The unfilled markers represent data instance from the train set. The filled markers represent data instance from the test set, and filled colour represents the class label assigned by SVM; whereas, the edge colour represents the true label. The marker size of the test set represents the probability that the sample instance is assigned with its corresponding class label; the bigger, the more confidence. Based on the obtained results shown in Table 3, the best prediction accuracy of the multi-class SVM model was 81.4%. It means that the overall classification accuracy is 81.4%, indicating the fact that 81.4 percent of injury severities related to individuals crashes was correctly identified. Moreover, the prediction accuracy of the two-class SVM model (85.3%) was significantly improved than multi-class classification (81.4%). When kernel function was polynomial, the accuracies of the SVM model were better than other kernel functions; meanwhile, the overall prediction accuracy from RBF and polynomial kernel functions was close to each other. LIBSVM had a better performance in terms of classification accuracy.

The MLP, which was applied for training, testing, and validation, consisted of 12 inputs, 25 neurons in the hidden layers, and 3 neurons in the output layer. The results of the MLP model are presented in Table 4 in the form of a prediction table. Table 4 depicts the prediction level of injury severity patterns in training, testing and validation phases.

The data for training, validation, and testing of the MLP application represented 70%, 15%, and 15% of all crash data, respectively. The overall accuracy in the MLP was 86.2%.

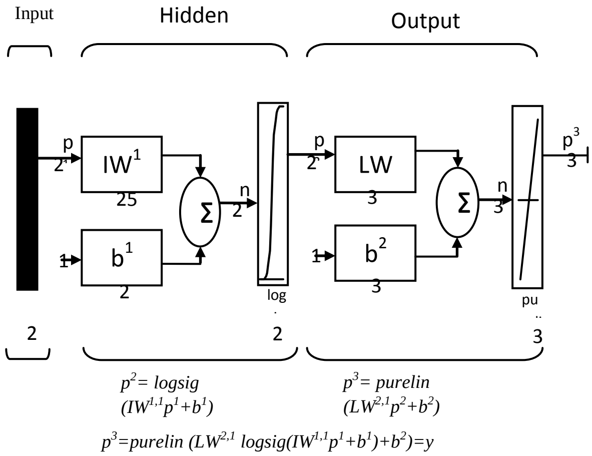
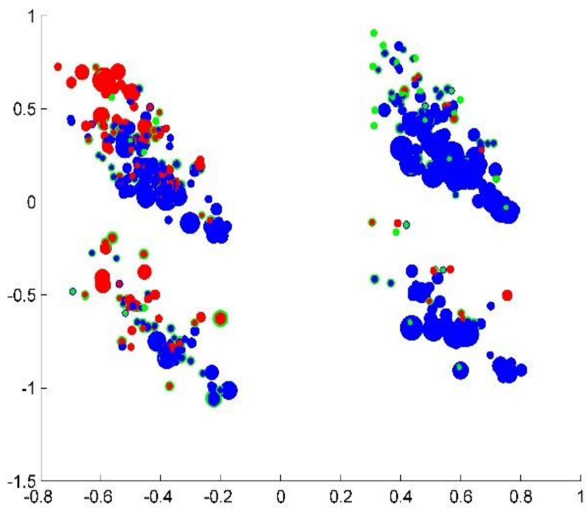
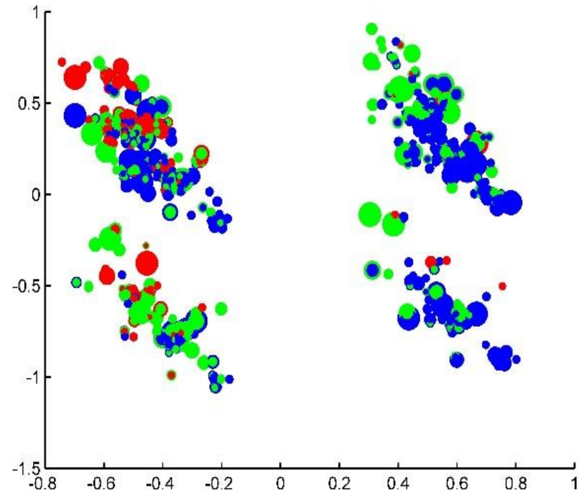


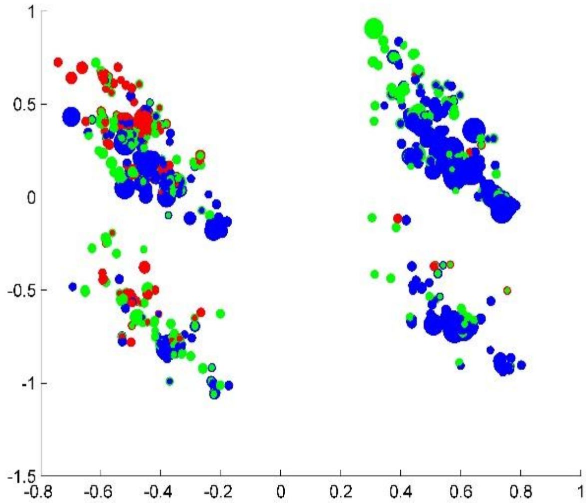
Figure 1. The structure of the final MLP neural



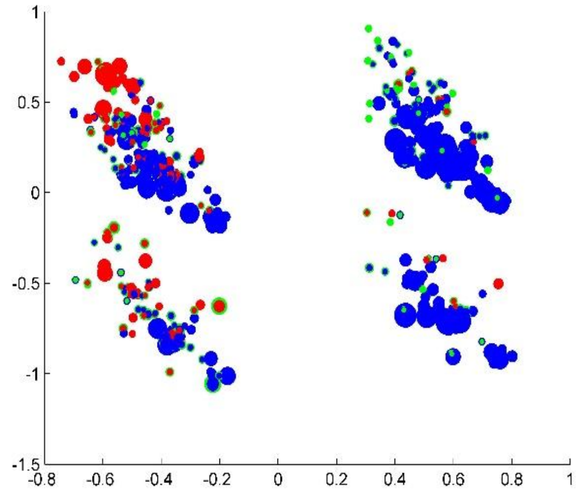
(a) Linear



(b) Polynomial



(c) RBF



(d) Sigmoid

Figure 2. Multi-class classification results of SVM for different kernel functions

Table 3. Classification accuracy of SVM models

Kernel function	LIBSVM		SVM^{light}	SVM^{multiclass}
	Avg. accuracy (two-class)	Avg. accuracy (multi-class)	Avg. accuracy (two- class)	Avg. accuracy (multi-class)
Linear	79.7	68.0	61.7	58.7
Polynomial	85.3	81.4	84.3	71.4
RBF	84.9	80.7	82.6	65.0
Sigmoid	79.8	68.6	62.1	58.9

Table 4. Prediction table of the MLP model

Injury severity	Training	Validation	Test	All
No Injury	89.8	80.9	82.6	87.9
Evident Injury	89.2	75.1	68.5	84.3
Fatality	88.4	68.5	74.7	82.5
Overall	90.1	77.6	76.4	86.2

Table 5 represents the overall prediction accuracy for the SVM, MLP, GA and combined GA and PS models. The results showed that the percent of prediction accuracy for the MLP model was 86.2%, which was higher than the SVM model with polynomial kernel (81.4%) followed by the combined GA and PS (78.6%) and GA (78.1%).

Table 5. Results of the SVM, MLP, GA and combined GA and PS models for crash injury severity

Result	SVM	MLP	GA	GA-PS
Overall Accuracy%	81.4	86.2	78.1	78.6

CONCLUSION

In this paper, the application of the SVM with the different kernel functions for crash injury severity prediction was investigated. The MLP, GA, combined GA and PS were also developed using the twelve input parameters and three levels of injury severity. The prediction accuracy of the SVM model was compared to the MLP, GA, combined GA and PS. The results demonstrated that the constructed MLP's performance was slightly better than the SVM for injury severity prediction. Whereas, SVM involved much lower computational cost than MLP because of using a straight forward algorithm in learning phase. The percent of prediction accuracy for the MLP mode was 86.2%, which was higher than the SVM model with polynomial kernel (81.4%) followed by the combination of the GA and PS (78.6%) and GA (78.1%). The classification results of the two-level SVM found to be 85.3% was higher than multi-class classification (81.4%). Also, the SVM prevented over fitting. Overall, using kernel functions led to calculating nonlinear solution, much simpler. One of the important advantage of SVM is to provide solutions with better generalization in comparison with MLP. A large number of parameters are necessary to be determined for MLP consisted of the number of hidden layers, number of hidden nodes, and transfer functions, etc. Meanwhile, few parameters are required for SVM according to the type of kernel functions. The advantage of using the GA or the combined GA and PS model is that the functions and coefficients of relationships are known. Thus, each model has its own advantage, and therefore using more than one method may provide a better understanding of the relationship between input and output variables.

REFERENCES

Abdel-Aty M, Pande A (2005). Identifying crash propensity using specific traffic speed conditions, Journal of Safety Research, 36(1): 97–108. doi:10.1016/j.jsr.2004.11.002

Abdelwahab HT, Abdel-Aty MA (2001). Development of artificial neural network models to predict driver injury severity in traffic accidents at signalized intersections, Transportation Research Record, 1746: 6–13. doi:10.3141/1746-02

Aghayan I, Noii N, Kunt MM (2012). Extended Traffic Crash Modeling through Precision and Response Time Using Fuzzy Clustering Algorithms Compared with Multi-Layer Perceptron. Promet- Traffic and Transportation, 24(6):455-467.

Aghayan I, Noii N, Kunt MM (2013). Fuzzy C-means clustering based on clustering algorithms for traffic crash data, Advances in Civil Engineering and Building Materials, Edited by Zhao, J., Taylor and Francis Group, London

Bedard M, Guyatt GH, Stones MJ, Hirdes JP (2002). The independent contribution of driver, crash, and vehicle characteristics to driver fatalities, Accident Analysis and Prevention, 34(6): 717–727. doi:10.1016/S0001-4575(01)00072-0

Ceylan H, Bell MGH (2004). Traffic signal timing optimization based on genetic algorithm approach, including drivers' routing, Transportation Research Part B: Methodological, 38(4): 329–342. doi:10.1016/S0191-2615(03)00015-8

Chang, CC, Lin, CJ (2013). A library for support vector machine,www.csie.ntu.edu.tw/~cjlin/libsvm/.

Chang, N.-B.; Chen, W. C. (2000). Prediction of PCDDs/

Cristianini N, Shawe-Taylor J (2000). An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, Cambridge University Press, Cambridge, UK.

Deschaine, LM, Francone F (2004). White Paper: Comparison of Discipulus™ (Linear Genetic Programming Software with Support Vector Machines, Classification Trees, Neural Networks and Human Experts). Available from Internet: <http://www.rmltech.com>.

Gun S (1998). Support vector machines for classification and regression, Technical Report, University of Southampton, UK.

Hadji Hosseinlou M, Aghayan I (2009). Prediction of traffic accident severity based on fuzzy logic, in 8th International Congress on Civil Engineering, 11–13 May 2009, Shiraz, Iran, 243–248.

Hensher DA, Ton TT (2000). A comparison of the predictive potential of artificial neural networks and nested logit models for commuter mode choice,

Transportation Research Part E: Logistics and Transportation Review, 36(3): 155–172. doi:10.1016/S1366-5545(99)00030-7
Holland, JH (1992). Adaptation in Natural and Artificial Systems: an Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence, A Bradford Book, 211 p.
Jin X, Cheu RL, Dipti S (2002). Development and adaptation of constructive probabilistic neural network in freeway incident detection, Transportation Research Part C: Emerging Technologies, 10(2): 121–147. doi:10.1016/S0968-090X(01)00007-9
Joachims T (1998). Making large-scale SVM learning practical. In B. Sch¨olkopf, C. J. C. Burges, and A. J. Smola (Eds.), Advances in Kernel Methods – Support Vector Learning, Cambridge, MA. MIT Press
Kunt MM, Aghayan I, Noii N (2011). Prediction for traffic accident severity: comparing the artificial neural network, genetic algorithm, combined genetic algorithm and pattern search methods. Transport, 26(4):353-366. doi:10.3846/16484142.2011.635465.
Li XG, Lord D, Zhang YL, Me YC (2008). Predicting motor vehicle crashes using support vector machine models, Accident Analysis and Prevention, 40(4):1611-1618.
Lord D, Manar A, Vizioli A. (2005). Modeling crash-flowdensity and crash-flow-V/C ratio relationships for rural and urban freeway segments, Accident Analysis and Prevention, 37(1): 185–199. doi:10.1016/j.aap.2004.07.003
Lv Y, Tang. S, Zhao H, Li S (2009). Real-time highway accident prediction based on support vector machines, Chinese Control and decision Conference, Chinese.4403-4407. doi:10.1109/CCDC.2009.5192409.
Mussone L, Ferrari A, Oneta M (1999). An analysis of urban collisions using an artificial intelligence model, Accident Analysis and Prevention, 31(6): 705–718. doi:10.1016/S0001-4575(99)00031-7
Mussone L, Rinelli S, Reitani G (1996). Estimating the accident probability of a vehicular flow by means of an artificial neural network, Environment and Planning B: Planning and Design, 23(6): 667–675. doi:10.1068/b230667
Ng, JCW, Sayed T (2004). Effect of geometric design consistency on road safety, Canadian Journal of Civil Engineering 31(2): 218–227, doi:10.1139/l03-090
Park B, Messer CJ, Urbanik II T (2000). Enhanced genetic algorithm for signal-timing optimization of oversaturated intersections, Transportation Research Record, 1727: 32–41. doi:10.3141/1727-05; PCDFs emissions from municipal incinerators by genetic programming and neural network modeling, Waste Management and Research, 18(4): 341–351. doi:10.1034/j.1399-3070.2000.00141.x
Sohn, SY, Lee SH (2003). Data fusion, ensemble and clustering to improve the classification accuracy for the severity of road traffic accidents in Korea, Safety Science, 41(1): 1–14. doi:10.1016/S0925-7535(01)00032-7
Subba Rao, PV, Sikdar PK, Krishna Rao KV, Dhingra, SL (1998). Another insight into artificial neural networks through behavioural analysis of access mode choice, Computers, Environment and Urban Systems, 22(5): 485–496. doi:10.1016/S0198-9715(98)00036-2
Teklu F, Sumalee A, Watling D. (2007). A genetic algorithm approach for optimizing traffic control signals considering routing, Computer-Aided Civil and Infrastructure Engineering, 22(1): 31–43. doi:10.1111/j.1467-8667.2006.00468.x
Tong HY, Hung WT (2002). Neural network modeling of vehicle discharge headway at signalized intersection: model descriptions and results, Transportation Research Part A: Policy and Practice, 36(1): 17–40. doi:10.1016/S0965-8564(00)00035-5
Valent F, Schiava, F, Savonitto C, Gallo T, Brusaferro, S, Barbone F. (2002). Risk factors for fatal road traffic accidents in Udine, Italy, Accident Analysis and Prevention, 34(1): 71–84. doi:10.1016/S0001-4575(00)00104-4
Vapnik, VN (1995). The Nature of Statistical Learning Theory, Springer Verlag, New York
Vythoulkas PC, Koutsopoulos HN (2003). Modeling discrete choice behavior using concepts from fuzzy set theory, approximate reasoning and neural networks, Transportation Research Part C: Emerging Technologies, 11(1): 51–73. doi:10.1016/S0968-090X(02)00021-9
Washington SP, Karlaftis MG, Mannering FL (2010). Statistical and Econometric Methods for Transportation Data Analysis, 2nd edition, Chapman and Hall/CRC, p.544
Wright CC, Abbess CR, Jarrett DF (1988). Estimating the regression-to-mean effect associated with road accident black spot treatment: towards a more realistic approach, Accident Analysis and Prevention, 20(3): 199–214. doi:10.1016/0001-4575(88)90004-8
Yin H, Wong SC, Xu J, Wong CK (2002). Urban traffic flow prediction using a fuzzy-neural approach, Transportation Research Part C: Emerging Technologies, 10(2): 85–98. doi:10.1016/S0968-090X(01)00004-3
Yuan F, Cheu RL (2003). Incident detection using support vector machines, Transportation Research Part C: Emerging Technologies, 11(3–4): 309–328. doi:10.1016/S0968-090X(03)00020-2
Zhang HM, Ritchie SG, Jayakrishnan R (2001). Coordinated traffic-responsive ramp control via nonlinear state feedback, Transportation Research Part C: Emerging Technologies, 9(5): 337–352. doi:10.1016/S0968-090X(00)00044-9
Zhang J, Lindsay J, Clarke K, Robbins G, Mao Y. (2000). Factors affecting the severity of motor vehicle traffic crashes involving elderly drivers in Ontario, Accident Analysis and Prevention, 32(1): 117–125. doi:10.1016/S0001-4575(99)00039-1
Zhong M, Lingras P, Sharma S. (2004). Estimation of missing traffic counts using factor, genetic, neural, and regression techniques, Transportation Research Part C: Emerging Technologies, 12(2): 139–166. doi:10.1016/j.trc.2004.07.006