Traffic Accident Data Mining Using Machine Learning Paradigms

Traffic Accident Data Mining Using Machine Learning Paradigms

Miao Chong, Ajith Abraham and Marcin Paprzycki
Computer Science Department,
Oklahoma State University, USA
ajith.abraham@ieee.org, marcin@cs.okstate.edu

Abstract **– Engineers and researchers in the automobile industry have tried to design and build safer automobiles, but traffic accidents are unavoidable. Patterns involved in dangerous crashes could be detected if we develop a prediction model that automatically classifies the type of injury severity of various traffic accidents. These behavioral and roadway patterns are useful in the development of traffic safety control policy. We believe that to obtain the greatest possible accident reduction effects with limited budgetary resources, it is important that measures be based on scientific and objective surveys of the causes of accidents and severity of injuries. This paper presents some models to predict the severity of injury that occurred during traffic accidents using three machine-learning approaches. We considered neural networks trained using hybrid learning approaches, decision trees and a concurrent hybrid model involving decision trees and neural networks. Experiment results reveal that among the machine learning paradigms considered the hybrid decision tree-neural network approach outperformed the individual approaches.**

I. INTRODUCTION AND RELATED RESEARCH

The costs of fatalities and injuries due to traffic accidents have a great impact on society. In recent years, researchers have paid increasing attention at determining the factors that significantly affect driver injury severity in traffic accidents.

Abdelwahab et al. studied the 1997 accident data for the Central Florida area [2]. The analysis focused on vehicle accidents that occurred at signalized intersections. The injury severity was divided into three classes: no injury, possible injury and disabling injury. They compared the performance of Multi-layered Perceptron (MLP) and Fuzzy ARTMAP, and found that MLP classification accuracy is higher than Fuzzy ARTMAP. Levenberg-Marquardt algorithm was used for MLP training and achieved 65.6 and 60.4 percent classification accuracy for the training and testing phases, respectively. Fuzzy ARTMAP achieved a classification accuracy of 56.1 percent.

Yang et al. used neural network approach to detect safer driving patterns that have less chances of causing death and injury when a car crash occurs [17]. They performed the Cramer's V Coefficient test [18] to identify significant variables that cause injury to reduce the dimensions of the data. Then, they applied data transformation method with a frequency-based scheme to transform categorical codes into numerical values. They used the Critical Analysis Reporting Environment (CARE) system, which was developed at the University of Alabama, using a Backpropagation (BP) neural network. They used the 1997 Alabama interstate alcohol-related data, and further studied the weights on the trained network to obtain a set of controllable cause variables that are likely causing the injury during a crash. The target variable in their study had two classes: injury and non-injury, in which injury class included fatalities. They found that by controlling a single variable (such as the driving speed, or the light conditions) they could reduce fatalities and injuries by up to 40%.

Sohn et al. applied data fusion, ensemble and clustering to improve the accuracy of individual classifiers for two categories of severity (bodily injury and property damage) of road traffic accident [15]. The individual classifiers used were neural network and decision tree. They applied a clustering algorithm to the dataset to divide it into subsets, and then used each subset of data to train the classifiers. They found that classification based on clustering works better if the variation in observations is relatively large as in Korean road traffic accident data.

Mussone et al. used neural networks to analyze vehicle accident that occurred at intersections in Milan, Italy [12]. They chose feed-forward MLP using BP learning. The model had 10 input nodes for eight variables (day or night, traffic flows circulating in the intersection, number of virtual conflict points, number of real conflict points, type of intersection, accident type, road surface condition, and weather conditions). The output node was called accident index, which was calculated as the ratio between the number of accidents for a given intersection and the number of accidents at the most dangerous intersection. Results showed that the highest accident index for running over of pedestrian occurs at non-signalized intersections at nighttime.

Dia et al. used real-world data for developing a multi-layered MLP neural network freeway incident detection model [5]. They compared the performance of the neural network model and the incident detection model in operation on Melbourne's freeways. Results showed that neural network model could provide faster and more reliable incident detection over the model that was in operation on Melbourne's freeways. They also found that failure to provide speed data at a station could significantly deteriorate model performance within that section of the freeway.

Shankar et al. applied a nested logic formulation for estimating accident severity likelihood conditioned on the occurrence of an accident [14]. They found that there is a greater probability of evident injury or disabling injury/fatality relative to no evident injury if at least one driver did not use a restraint system at the time of the accident.

Kim et al. developed a log-linear model to clarify the role of driver characteristics and behaviors in the causal sequence leading to more severe injuries. They found that alcohol or drug use and lack of seat belt use greatly increase the odds of more severe crashes and injuries [8].

Abdel-Aty et al. used the Fatality Analysis Reporting System (FARS) crash databases covering the period of 1975-2000 to analyze the effect of the increasing number of Light Truck Vehicle (LTV) registrations on fatal angle collision trends in the US [1]. They investigated the number of annual fatalities that resulted from angle collisions as well as collision configuration (car-car, car-LTV, LTV-car, and LTV-LTV). Time series modeling results showed that fatalities in angle collisions will increase in the next 10 years, and that they are affected by the expected increase in the percentage of LTVs in traffic.

Bedard et al. applied a multivariate logistic regression to determine the independent contribution of driver, crash, and vehicle characteristics to drivers' fatality risk [3]. They found that increasing seatbelt use, reducing speed, and reducing the number and severity of driver-side impacts might prevent fatalities.

Evanco conducted a multivariate population-based statistical analysis to determine the relationship between fatalities and accident notification times [6]. The analysis demonstrated that accident notification time is an important determinant of the number of fatalities for accidents on rural roadways.

Ossiander et al. used Poisson regression to analyze the association between the fatal crash rate (fatal crashes per vehicle mile traveled) and the speed limit increase [13]. They found that the speed limit increase was associated with a higher fatal crash rate and more deaths on freeways in Washington State.

Furthermore, some researchers studied the relationship between drivers' age, gender, vehicle mass, impact speed or driving speed measure with fatalities [4, 9, 10, 11, 16].

This paper investigates application of neural networks, decision trees and a hybrid combination of decision tree and neural network to build models that could predict injury severity. We also briefly report on our unsuccessful attempt at applying support vector machines to the problem. The remaining parts of the paper are organized as follows. In Section 2, more details about the problem and the pre-processing are presented, followed, in Section 3, by a short description the different machine learning paradigms used. Performance analysis is presented in Section 4 and finally some discussions and conclusions are given towards the end.

II. Accident Data Set

A. Description of Dataset

This study used data from the National Automotive Sampling System (NASS) General Estimates System (GES) [21]. The GES datasets are intended to be a nationally representative probability sample from the annual estimated 6.4 million accident reports in the United States. The dataset for the study contains traffic accident records from 1995 to 2000, a total number of 417,670 cases. According to the variable definitions for the GES dataset, this dataset has drivers' only records and doesn't include passengers' information. It includes labels of year, month, region, primary sampling unit, the number describing the police jurisdiction, case number, person number, vehicle number, vehicle make and model; inputs of drivers' age, gender, alcohol usage, restraint system, eject, vehicle body type, vehicle age, vehicle role, initial point of impact, manner of collision, rollover, roadway surface condition, light condition, travel speed, speed limit and the output injury severity. The injury severity has five classes: No Injury, Possible Injury, Non-incapacitating Injury, Incapacitating Injury, and Fatal Injury. In the original dataset, 70.18% of the cases have output of no injury, 16.07% of the cases have output of possible injury, 9.48% of the cases have output of non-incapacitating injury, 4.02% of the cases have output of incapacitating injury, and 0.25% of the cases have fatal injury.

Our task was to develop machine learning based intelligent models that could classify the severity of injuries (5 categories) more accurately. This can in turn lead to greater understanding of the relationship between the factors of driver, vehicle, roadway, and environment and driver injury severity. Accurate results of such data analysis can provide crucial information for the road accident prevention policy.

B. Data Preparation

The input and output variables are considered for the model building. There are no conflicts between the attributes since each variable represents its own characteristic. The variables are already categorized and are represented by numbers. The manner in which the collision occurred has 7 categories: not collision, rear-end, head-on, rear-to-rear, angle, sideswipe same direction, and sideswipe opposite direction. For these 7 categories the distribution of the fatal injury is as follows: 0.56% for not collision, 0.08% for rear-end collision, 1.54% for head-on collision, 0.00% for rear-to-rear collision, 0.20% for angle collision, 0.08% for sideswipe same direction collision, 0.49% for sideswipe opposite direction collision. Since *head-on* collision has the highest percent of fatal injury; therefore, the dataset was narrowed down to head-on collision only. Head-on collision has a total of 10,386 records. There are 160 records of head-on collision with fatal injury; all of these 160 records have the initial point of impact categorized as *front*.

The initial point of impact has 9 categories: no damage/non-collision, front, right side, left side, back, front right corner, front left corner, back right corner, back left corner. The head-on collision with *front impact* has 10,251 records; this is 98.70% of the 10,386 head-on collision records. We have therefore decided to focus on front impact only and removed the remaining 135 records. Travel speed and speed limit will not be used in the model because there are too many records with unknown value; for 67.68% of records the travel speed during accident and local speed limit were unknown. This means that the remaining input variables are: drivers' age, gender, alcohol usage, restraint system, eject, vehicle body type, vehicle role, vehicle age, rollover, road surface condition, light condition.

Vehicle age with values 37, 41, 46 and 56 each has only one record and these were the only records representing such old cars. These four records were therefore deleted from the dataset since they were clear outliers. Thus, finally, the dataset for modeling had 10,247 records. There were 5,171 (50.46%) records with no injury, 2138 (20.86%) records with possible injury, 1721 (16.80%) records with non-incapacitating injury, 1057 (10.32%) records with incapacitating injury, and 160 (1.56%) records with fatal injury. We have separated each output class and used one-against-all approach. This approach selects one output class to be the positive class, and all the other classes are combined to be the negative class. We set the output value of the positive class to 1, and the (combined) negative class(es) to 0. We divided the datasets randomly into 60%, 20%, and 20% for training, cross-validation, and testing respectively.

III. MACHINE LEARINING PARADIGMS

A. Hybrid Learning Artificial Neural Networks

Multilayer perceptron with backpropagation training is one of the standard neural network architectures. Basically, BP is a gradient descent technique to minimize the error E for a particular training pattern. For adjusting the weight (w_{ij}) from the i-th input unit to the j-th output, in the batched mode variant, the descent is based on the gradient ∇E ($\frac{\delta E}{\delta w_{ij}}$) for the total training set:

$$\Delta w_{ij}(n) = -\varepsilon * \frac{\delta E}{\delta w_{ij}} + \alpha * \Delta w_{ij}(n-1) \qquad (1)$$

The gradient gives the direction of the error E. The parameters ε and α are the learning rate and the momentum respectively. A good choice of both the parameters is required for training success and speed. Unfortunately there does not exist a practical approach that would allow successful automatic selection of best values of these parameters for a given dataset. Similarly there exist only few basic approaches to the selection of the optimal number of nodes in the hidden layer. We had therefore to establish these parameters experimentally. Empirical research [19] has shown that the BP used for training neural networks has the following problems:

- BP often gets trapped in a local minimum mainly because of the random initialization of weights.
- BP usually generalizes quite well to detect the global features of the input but after prolonged training the network will start to recognize individual input/output pair rather than settling for weights that generally describe the mapping for the whole training set.

The second popular training algorithm for neural networks is Scaled Conjugate Gradient Algorithm (SCGA). Moller [20] introduced it as a way of avoiding the complicated line search procedure of conventional conjugate gradient algorithm (CGA). According to the SCGA, the Hessian matrix is approximated by

$$E''(w_k)p_k = \frac{E'(w_k + \sigma_k p_k) - E'(w_k)}{\sigma_k} + \lambda_k p_k \qquad (2)$$

where E' and E'' are the first and second derivative information of global error function E (w_k). The other terms p_k, σ_k and λ_k represent the weights, search direction, parameter controlling the change in weight for second derivative approximation and parameter for regulating the indefiniteness of the Hessian. In order to get a good quadratic approximation of E, a mechanism to raise and lower λ_k is needed when the Hessian is positive definite. Detailed step-by-step description can be found in [20].

In order to minimize the above-mentioned problems due to BP training, we used a combination of BP and SCG for training.

B. Decision Trees

The decision tree model consists of a hierarchy of univariate binary decisions [7]. Each internal node in the tree specifies a binary test on a single variable, branch represents an outcome of the test, each leaf node represent class labels or class distribution. The decision tree algorithm operates by choosing the best variable for splitting the data into two groups at the root node, partitioning the data into two disjoint branches in such a way that the class labels in each branch are as homogeneous as possible, and splitting is recursively applied to each branch, and so forth. Once a maximal tree is generated, it examines smaller trees obtained by pruning away branches of the maximal tree. Once the maximal tree is grown and a set of sub-trees is derived from it, the decision tree algorithm determines the best tree by testing for misclassification error rates.

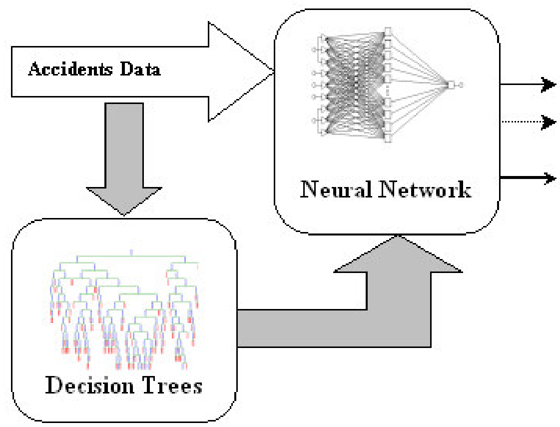

Fig. 1. Hybrid concurrent decision tree-ANN model for accident data

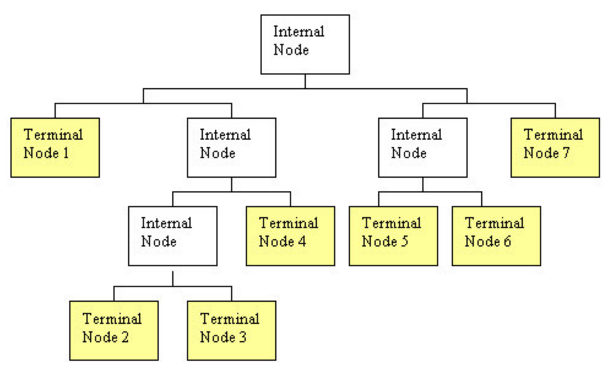

Fig. 2. Decision tree structure

C. Support Vector Machines (SVM)

SVMs are kernel-based learning algorithms in which only a fraction of the training examples are used in the solution (these are called the support vectors), and where the objective of learning is to maximize a margin around the decision surface. The flexibility of kernel functions allows the SVM to search a wide variety of hypothesis spaces. The basic idea of applying SVMs to pattern classification can be stated briefly as: first map the input vectors into one feature space (possibly with a higher dimension), either linearly or nonlinearly, which is relevant with the selection of the kernel function; then within the feature space, seek an optimized linear division, i.e. construct a hyperplane which separates two classes.

D. Hybrid Decision Tree-ANN (DTANN)

Figure 1 illustrates the hybrid decision tree-ANN (DTANN) model for predicting drivers' injury severity. We used a concurrent hybrid model where traffic accidents data are fed to the decision tree to generate the node information. Terminal nodes are numbered left to right starting with 1. All the data set records are assigned to one of the terminal nodes, which represent the particular class or subset. The training data together with the node information were supplied for training ANN. Figure 2 illustrates a decision tree structure. For the hybrid decision tree–ANN, we used the same hybrid learning algorithms and parameters setting as we used for ANN (except for the number of hidden neurons). Experiments were performed with different number of hidden neurons and models were selected with the highest classification accuracy for the output class.

IV. PERFORMANCE ANALYSIS

A. Neural Networks

Hyperbolic activation function was used in the hidden layer and logistic activation function in the output layer. The models were trained with BP (100 epochs, learning rate 0.01) and SCGA (500 epochs) to minimize the Mean Squared Error (MSE). For each output class, we experimented with different number of hidden neurons, and selected the model with highest classification accuracy for the class. From the results of experiments, for no injury class the best model had 65 hidden neurons, and achieved training and testing performance of 63.86% and 60.45% respectively. For possible injury class, the best model had 65 hidden neurons achieving it's training and testing performance of 59.34% and 57.58% respectively. For non-incapacitating injury class, the best model had 75 hidden neurons achieving training and testing performance of 58.71% and 56.8% respectively. For incapacitating injury class, the best model had 60 hidden neurons achieving training and testing performance of 63.40% and 63.36% respectively. Finally, for fatal injury class, the best model had 45 hidden neurons achieving training and testing performance of 78.61% and 78.17% respectively. These results are the summary of multiple experiments as illustrated in Table 2.

B. Decision Trees

We trained each class with Gini goodness of fit measure, the prior class probabilities was set to equal, the stopping option for pruning was misclassification error, the minimum n per node was set to 5, fraction of objects was 0.05, the maximum number of nodes was 1000, the maximum number of level in tree was 32, the number of surrogates was 5, we used 10 fold cross-validation, and generated the comprehensive results. The cross-validation testing ensures that the patterns found will hold up when applied to new data. The performances for no injury, possible injury, non-incapacitating injury, incapacitating injury and fatal injury were 67.54%, 64.39%, 60.37%, 71.38%, and 89.46% respectively. Empirical results including classification matrix are illustrated in Table 1.

C. Support Vector Machines

We experimented with polynomial kernel and radial basis function kernel. For some reason, polynomial kernel was not that successful and hence we only focused on radial basis function (RBF) kernels. Table 3 illustrates the SVM performance for the different parameter settings and the accuracy of each experiment for each class.

TABLE 1. DECISION TREE PERFORMANCE

Injury Class	DT Accuracy (%)
No Injury:	67.54
Possible Injury:	64.40
Non-incapacitating Injury:	60.37
Incapacitating Injury:	71.38
Fatal Injury:	89.46

D. Hybrid DT-ANN Approach

From the experiment results, for no injury class the best model had 70 hidden neurons, with training and testing performance of 83.02% and 65.12% respectively. For possible injury class, the best model had 98 hidden neurons with training and testing performance of 74.93% and 63.10% respectively. For non-incapacitating injury class, the best model had 109 hidden neurons with training and testing performance of 71.88% and 62.24% respectively. For incapacitating injury class, the best model had 102 hidden neurons, with training and testing performance of 77.95% and 72.63% respectively. For fatal injury class, the best model had 76 hidden neurons with training and testing performance of 91.53% and 90.00% respectively. These are the best models out of multiple experiments. Empirical results are presented in Table 4 and a final comparison between ANN, DT and DTANN is graphically illustrated in Figure 3. For all the output classes, the hybrid DTANN outperformed the ANN. For non-incapacitating injury, incapacitating injury, and fatal injury classes, the hybrid DTANN outperformed both ANN and DT.

TABLE 2. NEURAL NETWORK PERFORMANCE

No Injury			Possible Injury			Non-incapacitating			Incapacitating			Fatal Injury		
# neurons	Accuracy %		# neurons	Accuracy %		# neurons	Accuracy %		# neurons	Accuracy %		# neurons	Accuracy %	
	Train	Test		Train	Test		Train	Test		Train	Test		Train	Test
60	63.57	59.67	65	59.34	**57.58**	60	57.88	55.25	60	63.4	**63.36**	45	77.26	**75.17**
65	63.86	**60.45**	70	59.56	55.15	65	57.69	54.66	65	62.23	61.32	57	74.78	70.65
70	63.93	60.25	75	58.88	57.29	75	58.71	**56.80**	75	61.06	61.52	65	69.81	69.73
75	64.38	57.43	80	58.39	56.22	80	57.78	54.13	84	63.23	58.41	75	60.19	59.62
80	63.64	58.89	95	60.07	55.93	85	57.83	55.59	90	59.32	59.08	80	74.33	71.77

TABLE 3: PERFORMANCE OF SVM USING RADIAL BASIS FUNCTION KERNEL

	g=0.0001 c=42.8758	g=0.001 c=4.6594	g=0.5 c=0.5	g=1.2 c=0.5	g=1.5 c=2	g=2 c=10	g=0.00001 c=100	g=0.0001 c=100	g=0.001 c=100
No injury									
Class 0	59.76	59.80	57.95	57.65	53.62	54.12	57.34	59.76	60.46
Class 1	60.14	60.14	60.82	55.63	55.73	55.53	62.88	60.14	60.14
Possible injury									
Class 0	100.00	100.00	100.00	99.88	95.33	95.58	100.00	100.00	100.00
Class 1	0.00	0.00	0.00	0.00	3.67	3.42	0.00	0.00	0.00
Non-incapacitating									
Class 0	100.00	100.00	100.00	100.00	97.43	97.49	100.00	100.00	100.00
Class 1	0.00	0.00	0.00	0.00	3.21	2.92	0.00	0.00	0.00
Incapacitating									
Class 0	100.00	100.00	100.00	99.89	98.06	98.11	100.00	100.00	100.00
Class 1	0.00	0.00	0.00	0.00	2.83	2.83	0.00	0.00	0.00
Fatal Injury									
Class 0	100.00	100.00	100.00	100.00	99.95	99.95	100.00	100.00	100.00
Class 1	0.00	0.00	0.00	0.00	3.33	3.33	0.00	0.00	0.00

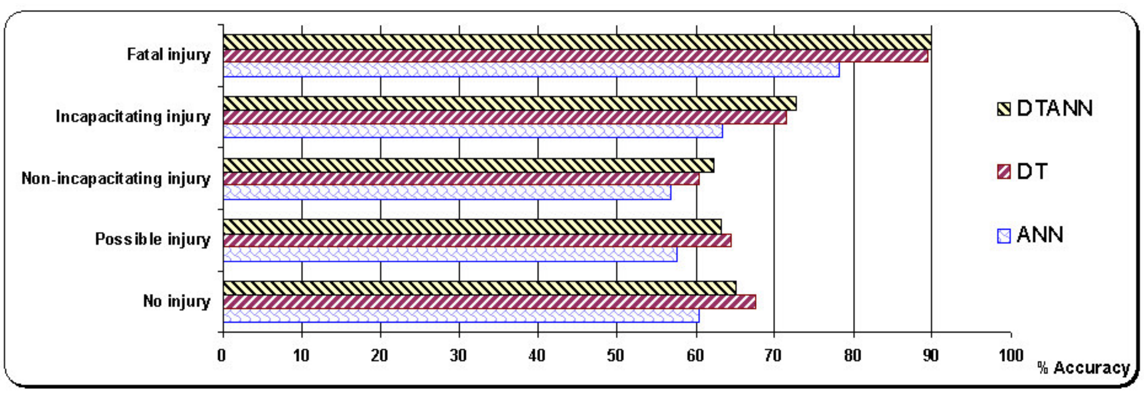

Fig. 3. Performance comparison of the different learning paradigms

TABLE 4. TEST PERFORMANCE OF DTANN

Injury type	% Accuracy
	DTANN
No Injury	65.12
Possible Injury	63.10
Non-Incapacitating Injury	62.24
Incapacitating Injury	72.63
Fatal Injury	90.00

V. CONCLUSIONS

In this paper, we studied the GES automobile accident data from 1995 to 2000 and investigated the performance of neural network, decision tree, support vector machines and a hybrid decision tree - neural network for predicting drivers' injury severity in head-on front impact point collisions. The classification accuracy on the test results reveals that, for non-incapacitating injury, incapacitating injury, and fatal injury classes, the hybrid approach performed better than neural network, decision trees and support vector machines. For no injury and possible injury classes, the hybrid approach performed better than neural network. The no injury and possible injury classes could be best modeled by decision trees.

Previous researches had focus mainly on no injury and injury (including fatality) classes. In this paper, we extended the research to possible injury, non-incapacitating injury, incapacitating injury, and fatal injury. Our experiments showed that the model for fatal and non-fatal injury performed better than other classes. The ability of predicting fatal and non-fatal injury is very important since drivers' fatality has the highest cost to society economically and socially.

One very important factor of causing different injury level is the actual speed that the vehicle was going when the accident happened. Our dataset doesn't provide enough information on the actual speed since speed for 67.68% of the data records' was unknown. If the speed was available, it might help to improve the models performance. From an intelligent systems point of view it is interesting to note about the failure of SVMs to model the complexity of the different injury classes.

VI. References

[1] Abdel-Aty, M., and Abdelwahab, H., Analysis and Prediction of Traffic Fatalities Resulting From Angle Collisions Including the Effect of Vehicles' Configuration and Compatibility. *Accident Analysis and Prevention*, 2003.

[2] Abdelwahab, H. T. and Abdel-Aty, M. A., Development of Artificial Neural Network Models to Predict Driver Injury Severity in Traffic Accidents at Signalized Intersections. *Transportation Research Record 1746*, Paper No. 01-2234.

[3] Bedard, M., Guyatt, G. H., Stones, M. J., & Hireds, J. P., The Independent Contribution of Driver, Crash, and Vehicle Characteristics to Driver Fatalities. *Accident analysis and Prevention*, Vol. 34, pp. 717-727, 2002.

[4] Buzeman, D. G., Viano, D. C., & Lovsund, P., Car Occupant Safety in Frontal Crashes: A Parameter Study of Vehicle Mass, Impact Speed, and Inherent Vehicle Protection. *Accident Analysis and Prevention*, Vol. 30, No. 6, pp. 713-722, 1998.

[5] Dia, H., & Rose, G., Development and Evaluation of Neural Network Freeway Incident Detection Models Using Field Data. *Transportation Research C*, Vol. 5, No. 5, 1997, pp. 313-331.

[6] Evanco, W. M., The Potential Impact of Rural Mayday Systems on Vehicular Crash Fatalities. Accident Analysis and Prevention, Vol. 31, 1999, pp. 455-462.

[7] Hand, D., Mannila, H., & Smyth, P., Principles of Data Mining. The MIT Press, 2001.

[8] Kim, K., Nitz, L., Richardson, J., & Li, L., Personal and Behavioral Predictors of Automobile Crash and Injury Severity. *Accident Analysis and Prevention*, Vol. 27, No. 4, 1995, pp. 469-481.

[9] Kweon, Y. J., & Kockelman, D. M., Overall Injury Risk to Different Drivers: Combining Exposure, Frequency, and Severity Models. Accident Analysis and Prevention, Vol. 35, 2003, pp. 441-450.

[10] Martin, P. G., Crandall, J. R., & Pilkey, W. D., Injury Trends of Passenger Car Drivers In the USA. Accident Analysis and Prevention, Vol. 32, 2000, pp. 541-557.

[11] Mayhew, D. R., Ferguson, S. A., Desmond, K. J., & Simpson, G. M., Trends In Fatal Crashes Involving Female Drivers, 1975-1998. Accident Analysis and Prevention, Vol. 35, 2003, pp. 407-415.

[12] Mussone, L., Ferrari, A., & Oneta, M., An analysis of urban collisions using an artificial intelligence model. *Accident Analysis and Prevention*, Vol. 31, 1999, pp. 705-718.

[13] Ossiander, E. M., & Cummings, P., Freeway speed limits and Traffic Fatalities in Washington State. Accident Analysis and Prevention, Vol. 34, 2002, pp. 13-18.

[14] Shankar, V., Mannering, F., & Barfield, W., Statistical Analysis of Accident Severity on Rural Freeways. *Accident Analysis and Prevention*, Vol. 28, No. 3, 1996, pp.391-401.

[15] Sohn, S. Y., & Lee, S. H., Data Fusion, Ensemble and Clustering to Improve the Classification Accuracy for the Severity of Road Traffic Accidents in Korea. *Safety Science*, Vol. 4, issue1, February 2003, pp. 1-14.

[16] Tavris, D. R., Kuhn, E. M, & Layde, P. M., Age and Gender Patterns In Motor Vehicle Crash injuries: Improtance of Type of Crash and Occupant Role. Accident Analysis and Prevention, Vol. 33, 2001, pp. 167-172.

[17] Yang, W.T., Chen, H. C., & Brown, D. B., Detecting Safer Driving Patterns By A Neural Network Approach. *ANNIE '99 for the Proceedings of Smart Engineering System Design Neural Network, Evolutionary Programming, Complex Systems and Data Mining*, Vol. 9, pp 839-844, Nov. 1999.

[18] Zembowicz, R. and Zytkow, J. M., 1996. From Contingency Tables to Various Forms of Knowledge in Database. Advances in knowledge Discovery and Data Mining, editors, Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R. AAAI Press/The MIT Press, pp.329-349.

[19] Abraham A., Meta-Learning Evolutionary Artificial Neural Networks, Neurocomputing Journal, Elsevier Science, Netherlands, Vol. 56c, pp. 1-38, 2004.

[20] Moller A.F., A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning, Neural Networks, Volume (6), pp. 525-533, 1993.

[21] National Center for Statistics and Analysis <http://www-nrd.nhtsa.dot.gov/departments/nrd-30/ncsa/NASS.html>