Paper No. 01-2234

Development of Artificial Neural Network Models to Predict Driver Injury Severity in Traffic Accidents at Signalized Intersections

Hassan T. Abdelwahab and Mohamed A. Abdel-Aty

The relationship between driver injury severity and driver, vehicle, roadway, and environment characteristics was examined. The use of two well-known neural network paradigms, the multilayer perceptron (MLP) and fuzzy adaptive resonance theory (ART) neural networks, was investigated. The use of artificial neural networks can lead to greater understanding of the relationship between the aforementioned factors and driver injury severity. Accident data for 1997 for the Central Florida area, which consists of Orange, Osceola, and Seminole Counties, were used. The analysis focuses on two-vehicle accidents that occurred at signalized intersections. The MLP neural network has a better generalization performance of 65.6 and 60.4 percent for the training and testing phases, respectively. The performance of the MLP was compared with that of an ordered logit model. The ordered logit model was able to correctly classify only 58.9 and 57.1 percent for the training and testing phases, respectively. A simulation experiment was then carried out to understand the MLP neural network model. Results show that rural intersections are more dangerous in terms of driver injury severity than urban intersections. Also, female drivers are more likely to experience a severe injury than are male drivers. Speed ratio increases the likelihood of injury severity. Drivers at fault are less likely to experience severe injury than are those not at fault. Wearing a seat belt decreases the chance of sustaining severe injuries. Vehicle type plays a role in driver injury severity. Drivers in passenger cars are more likely to experience a greater injury severity level than are drivers of vans or pickup trucks. Finally, drivers exposed to impact at their side experience greater injury severity than those exposed to impact elsewhere.

In recent years increased attention has been directed at determining the factors that significantly affect driver injury severity in traffic accidents. Several approaches have been employed to study injury severity.

Lui et al. used a logistic regression approach to model the probability of fatalities conditioned on the occurrence of an accident (*1*). However, the analysis was limited to two-vehicle accidents with at least one death. The probability of fatality was modeled as a function of the driver age and gender, impact points, car deformation, driver safety belt status, and car weight.

Nasser et al. studied the accident severity on a microlevel approach based on a sequential logit model (*2*). Three accident types were considered: single-vehicle accident, two-vehicle accident, and multivehicle accident. Factors found to be significant include an accident dynamic term (a combination of vehicle speed, vehicle mass, and direction of impact), seat belt use, alcohol, driver condition, and driver fault. Road surface condition was insignificant in the models. Kim et al. used accident type, seat belt use, and injury severity variables to find the relationship among these three factors (*3*). Kim et al. estimated a log-linear model to investigate the role of driver characteristics and behaviors in the causal sequence leading to more-severe injuries (*4*). They found that the driver behaviors of alcohol or drug use and lack of seat belt use greatly increase the odds of more severe accidents and injuries.

Shankar et al. applied a two-level nesting logit formulation to predict the accident severity given that an accident has occurred (*5*). Results show a greater probability of evident injury or fatality relative to no evident injury if at least one driver did not use a restraint system at the time of the accident. Edwards investigated the relationship between recorded weather and road accidents in England and Wales (*6*). Accident severities for various adverse weather categories of rain, fog, and high winds were compared with that for the nonhazardous condition of fine weather. Findings showed that accident severity decreases significantly in rain as opposed to fine weather.

Artificial neural networks (ANNs) have been successfully applied in several transportation problems (*7–10*). Mussone et al. applied ANNs to analyze vehicular accidents that occurred at intersections in Milan, Italy (*11*). The model has 10 input nodes, 4 hidden nodes, and 1 output node. The input nodes represent the following variables: day or night, traffic flow, number of conflict points, type of intersection, accident type, road surface condition, and weather conditions. The output node was a point value called accident index. An accident index 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 neural network models are capable of extracting information in terms of the factors that explain accidents and the factors contributing to a greater degree of danger.

The major objective of this study was to investigate the use of ANNs in predicting injury severity at signalized intersections given that an accident has occurred. Various ANN architectures were investigated to predict driver injury severity. These neural network models can improve the understanding of the relationship between driver injury severity and other factors related to the driver, vehicle, and roadway/environment. Three levels of severity were considered: no injury, possible/evident injury (minor injury), and disabling injury/fatality. Driver characteristics are age, gender, alcohol usage (driving under the influence or not), fault (at fault or not), seat belt usage, and speed. Vehicle characteristics are vehicle type and point of impact. Roadway/environment conditions are area type (rural versus urban), day (weekend versus weekday), time (off peak versus peak), light condition (daylight versus night), and weather condition (clear versus not clear).

University of Central Florida, Department of Civil and Environmental Engineering, Orlando, FL 32816.

ANNs

Driver injury severity can be classified as belonging to one of three discrete categories: no injury, possible/evident injury, and disabling injury/fatality. Therefore, an injury severity model can be addressed as a pattern recognition task that can best be solved using ANNs. The problem is to map an input vector of accident-related characteristics into an output space of dimensionality 3 (number of severity categories). All drivers have the same output space (i.e., any driver can be involved in an accident with one of the three severity levels). An ANN is a network of many simple processors (units, nodes, or neurons), each one of which has a small amount of local memory. These processors are connected by unidirectional communication channels (connections), which carry numeral data. An ANN resembles the brain in two aspects: (*a*) knowledge is acquired by the network through a learning process and (*b*) interneuron connection strengths known as synaptic weights are used to store the knowledge. ANNs offer the following salient characteristics and properties (*12*):

1. Nonlinear input-output mapping: ANNs can learn arbitrary nonlinear input-output mapping directly from training data.
2. Generalization: ANNs can sensibly interpolate input patterns that are new to the network. From a statistical point of view, ANNs can fit the desired function in such a way that they have the ability to generalize to situations that differ from those in the collected training data.
3. Adaptivity: ANNs can automatically adjust their connection weights, or even network structure (number of nodes or connection types) to optimize their behavior as controllers, predictors, pattern recognizers, decision makers, and so forth.
4. Fault tolerance: The inherent fault-tolerance capability of ANNs stems from the fact that the large number of connections provides much redundancy, that each node acts independently of all others, and that each node relies on local information.

ANNs Versus Traditional Statistical Models

There has been an increasing interest in the applicability of ANNs in disparate domains. For example, with regard to the statistical regression methods, a principal strength of the ANN approaches is that they are explicitly nonlinear through hidden layers. Because they involve a more general mapping procedure, a specific function format is not required in model building. To adopt the regression method effectively, the form of the fitting function should be defined in advance. A fitting should then be carried out according to the minimal sum of squared errors. However, if the form of the fitting function cannot be specified, then the ANN-based method is suggested. It does not require any a priori information and can map the input patterns to the output patterns properly.

Some studies have compared several statistical methods with ANNs. Subba et al. (*9*) presented a comparative evaluation between the multinomial logit model (MNL) and ANN in modeling the choice behavior with respect to the access mode for transit. Because of highly nonlinear internal relationships, which are established by the ANN, its performance has been found to be much superior to that of MNL in both calibration and prediction.

Hashemi et al. developed an ANN model to predict vessel accidents on the lower Mississippi River (*13*). The predictive capability for vessel accidents of an ANN was compared with multiple discriminant analysis and logistic regression. Ninety percent of cases were correctly classified by this network, whereas logistic regression correctly classified only 56 percent.

Nijkamp et al. presented an analysis of a modal split problem using a traditional logit model and a feed-forward ANN model (*14*). The results highlighted the fact that the two models, although methodologically different, are both able to provide a reasonable forecasting of the problem studied. In particular, the ANN model turned out to have a slightly better performance.

Multilayer Perceptron Neural Networks

The multilayer perceptron (MLP) usually consists of three layers, as shown in Figure 1: input layer, hidden layer, and output layer. In this figure, input layer has *K* nodes and a bias node (designated as Node 0), the hidden layer has *J* nodes and a bias node (designated as Node 0), and the output layer has *I* nodes and no bias node. The connections in the MLP are of the feed-forward type. That is, connections are allowed from a layer of a certain index to layers of a higher index. As can be seen from Figure 1, no connections are allowed from a layer of a certain index to a layer of lower index (i.e., feedback connections). Also, no connections are allowed among the nodes belonging to the same layer.

The MLP neural networks can operate in two distinct phases: the training phase and the testing phase. The training phase works as follows: given a collection of training data $\{x(1), d(1)\}, \ldots, \{x(p), d(p)\}, \ldots, \{x(PT), d(PT)\}$, the objective is to map $\{x(1)$ to $d(1)\}, \ldots, \{x(p)$ to $d(p)\}$, and eventually $\{x(PT)$ to $d(PT)\}$. The backpropagation learning algorithm (*15*) is used to train the MLP. Figure 2 illustrates procedures of a simple backpropagation algorithm.

Consider that an input pattern $x(p)$ applied across the input layer of the MLP. Then, to ensure that the output of the MLP is equal to $d(p)$, an error function of the following form is constructed:

$$E(\mathbf{w}) = \sum_{p=1}^{PT} \sum_{i=1}^{I} \left[d_i^2(p) - y_i^2(p) \right]^2 \tag{1}$$

where

$E(\mathbf{w})$ = error function to be minimized,
$\mathbf{w}$ = weight vector,
PT = number of training patterns,
I = number of output nodes,
$d_i^2(p)$ = desired output of node i when pattern p is introduced to the MLP, and
$y_i^2(p)$ = actual output of node i when pattern p is introduced to the MLP.

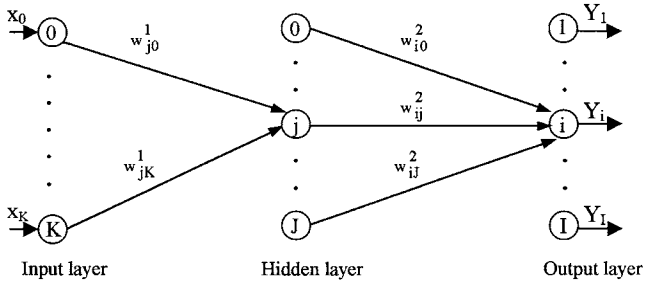

FIGURE 1 MLP neural network architecture.

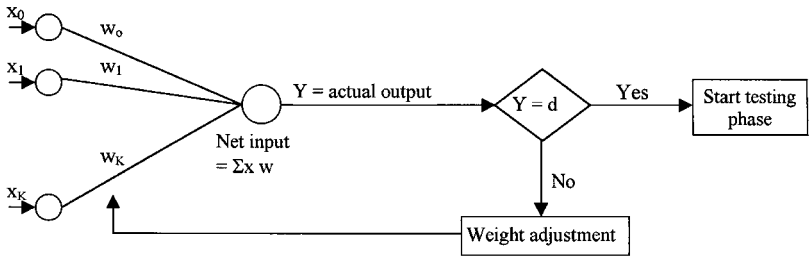

FIGURE 2 Simple backpropagation algorithm.

The objective is to change the weight vector **w** so that the above function is minimized. By minimizing the above error function, the actual output is in fact driven closer to the desired output. Assuming that the aforementioned error function is differentiable, well-known optimization procedures can be applied to minimize it. One of these procedures is the gradient descent procedure, which changes the weight vector **w** by an amount that is proportional to the negative gradient. That is,

$$\Delta \mathbf{w} = -\eta \nabla E(\mathbf{w}) \tag{2}$$

where

$\Delta \mathbf{w}$ = change of weight vector,
η = learning parameter, and
$\nabla E(\mathbf{w})$ = gradient vector $E(\mathbf{w})$ with respect to weight vector **w**.

It can be shown that the equations that describe the change of the weights are as follows:

$$\Delta \mathbf{w}_{ij}^{2} = \eta \delta_i^2(p) y_j^1(p) \tag{3}$$

$$\Delta \mathbf{w}_{jk}^{1} = \eta \delta_j^1(p) x_k(p) \tag{4}$$

where

$\Delta \mathbf{w}_{ij}^{2}$ = weight converging to the output layer,
$\Delta \mathbf{w}_{jk}^{1}$ = weight converging to the hidden layer,
$\delta_i^2(p)$ = error term associated with output node i due to presentation of input pattern p,
$\delta_j^1(p)$ = error term associated with hidden node j due to presentation of input pattern p,
$y_j^1(p)$ = output of hidden node j due to presentation of input pattern p, and
$x_k(p)$ = input component of index k in input pattern p.

Fuzzy ARTMAP Neural Networks

Fuzzy ARTMAP is a member of the class of ANN architectures referred to as ART architectures developed by Carpenter et al. (*16*). The ART architectures are based on the adaptive resonance theory (ART), which was introduced by Grossberg (*17*). Fuzzy ARTMAP is a clustering algorithm that maps a set of input vectors to a set of clusters. This algorithm has advantages over some other types of ANNs. A fuzzy ARTMAP model provides fast, stable learning in response to analog or binary input patterns. The fuzzy ARTMAP model was selected since it is the most recent model in the ART family. The fuzzy ARTMAP neural networks consist of two fuzzy ART modules, designated as ART_a and ART_b, as well as an inter-ART module (*18*). Inputs are presented at the ART_a module, and their corresponding outputs are presented at the ART_b module. The inter-ART module includes a MAP field whose purpose is to determine whether the correct mapping has been established.

A detailed description of the fuzzy ARTMAP ANN is given by Carpenter et al. (*16*). This section introduces only a brief description of the fuzzy ARTMAP training algorithm. Before the start of training, some preprocessing operations of the input patterns take place before they are presented to the ART_a module. The first preprocessing stage takes an input of M_a-dimensional input pattern from the pattern classification task and transforms it into an output vector $\mathbf{a} = (a_1, \ldots, a_{Ma})$, whose every component lies in the interval [0, 1]. The second preprocessing stage accepts as input the vector **a** of the first preprocessing stage and produces a vector **I**, such that

$$\boldsymbol{I} = \left(\boldsymbol{a}, \boldsymbol{a}^c\right) = \left(a_1, \ldots, a_{Ma}, a_1^c, \ldots, a_{Ma}^c\right) \tag{5}$$

where $a_i \in [0,1]$ and $a_i^c = 1 - a_i$ with $1 \le i \le M_a$.

The above transformation is called complement coding. These normalization and complementing processes are required by the fuzzy ARTMAP training algorithm. Similar types of operations are also performed to produce the output pattern that is applied at the ART_b module. The output pattern, designated by **O**, has the form

$$\mathbf{O} = \left(b_1, \ldots, b_{M_b}, b_i^c, \ldots, b_{M_b}^c\right) \tag{6}$$

where $b_i \in [0,1]$ and $b_i^c = 1 - b_i$ with $1 \le i \le M_b$.

The training phase of fuzzy ARTMAP works as follows. Given a list of training input/output pairs, such as $\{I^1, O^1\}, \ldots \{I^r, O^r\}, \ldots \{I^{PT}, O^{PT}\}$, the objective is to train fuzzy ARTMAP to map every input pattern of the training list to its corresponding output pattern. To achieve this goal, the training list is repeatedly presented to the fuzzy ARTMAP architecture in this order: I^1 to ART_a and O^1 to ART_b, then I^2 to ART_a and O^2 to ART_b, and finally I^{PT} to ART_a and O^{PT} to ART_b. This corresponds to one list presentation. The training list is presented as many times as it is necessary for fuzzy ARTMAP to correctly classify all the input patterns. The task is considered accomplished (i.e., the learning is complete) when the weights do not change during a list presentation. The performance phase occurs when the trained fuzzy ARTMAP network is used to classify a list of test input patterns.

It has been documented in the literature that the performance of fuzzy ARTMAP depends on the values of two parameters called the choice and vigilance parameters and on the order of pattern presentation (*18*). The dependence of fuzzy ARTMAP on the choice and vigilance parameters is an inherent characteristic of the algorithm. This vigilance parameter measures how close the input pattern must be to the top-down prototype for resonance to occur. This parameter has a range of 0 to 1. Low values of ρ_a ($\rho_a \approx 0$) lead to

broad generalization and coarse clustering, while high values ($\rho_a \approx 1$) lead to narrow generalization and fine clustering. Most fuzzy ARTMAP simulations that have appeared in the literature assume 0 values for the choice and vigilance parameters. A primary reason for the popularity of this choice is that it tends to minimize the size of the resulting network architecture.

If the choice and vigilance parameters are both set to 0, one ends up with a fuzzy ARTMAP algorithm that exhibits a significant variation in generalization performance for different orders of training pattern presentations. Because the performance of fuzzy ARTMAP depends on the order of pattern presentation in the training set, different orders of pattern presentation should be tried so as to achieve the best generalization performance. However, it is not an easy task to guess which one of the exceedingly large number of orders of pattern presentation exhibits the best generalization. In this paper, the training data were preprocessed by applying a systematic clustering procedure [the K-means algorithm (*19*)], to identify a fixed order of presenting the training input pattern. Herein, this procedure is referred to as the ordering algorithm, and fuzzy ARTMAP as ordered fuzzy ARTMAP (O-ARTMAP). The K-means ordering algorithm is based on the minimization of a performance index, which is defined as the sum of the squared Euclidean distances from all patterns in a cluster domain to the cluster center.

DATA

The 1997 accident data for the Central Florida area were used in this study. The Central Florida area consists of three counties: Orange, Osceola, and Seminole. The data were obtained from the 1997 Florida accident databases. The files used in this study are events, vehicle, and driver. The events file contains general information about the accident characteristics and circumstances. The vehicle file contains information about the vehicles and vehicle maneuver before the accident. The driver file contains information about the drivers and the condition or action of the driver that contributed to the accident. The field that contains a report number (a number that uniquely identifies each accident) was used to link the data among the three files.

The analysis focuses on two-vehicle accidents that occurred at signalized intersections. A total of 2,336 drivers were involved in 1,168 traffic accidents reported during 1997; 1,088 (46.6 percent) driver involvements resulted in no injury, 1,108 (47.4 percent) in possible/evident injury, and 140 (6.0 percent) in disabling injury. Table 1 shows a summary of injury severity distribution by several driver characteristics, vehicle factors, and road/environmental conditions. As seen in Table 1, young drivers aged 25 to 34 have a higher proportion of traffic accidents causing no injury than do drivers of other ages. Female drivers tend to be involved in a higher proportion of accidents causing possible or evident injury and disabling injury.

The use of seat belts has a major effect on driver injury severity. Drivers not wearing seat belts at the time of an accident were involved in 14.7 percent of the disabling or fatal injury accidents. Drivers in passenger cars experience a higher percentage of disabling or fatal injuries. Speed ratio, defined as the ratio of the estimated speed at the time of the accident to the posted speed limit, has an effect on the injury severity; higher speed ratios tend to result in higher injury severity. The point of impact has a great effect on the driver injury severity; accidents that occur at the driver's side tend to produce a higher injury severity level.

Categorical data analysis and the ANNs were used to identify the set of the candidate variables. The ANN modeling approach used in this study started with a full model that contains all the independent variables (13 variables as listed in Table 1). To assess the significance of each input variable (or a set of input variables) of the input vector, a model was constructed without that variable (or variables). Performance of the reduced model was then compared with the full model. Categorical data analysis techniques (*20*) were also used to test the significance of the two-way associations shown in Table 1. Driver injury severity is ordinal in nature. Some of the accident factors are ordinal (e.g., age and speed ratio), whereas others are nominal. Therefore, the statistical measures have to vary according to the factor type. If variables (X and Y) in a two-way table are ordinal, the pair of subjects can be classified as concordant or discordant. The pair is concordant if the subject ranking higher on variable X also ranks higher on variable Y. The pair is discordant if the subject ranking higher on X ranks lower on Y. Then the statistic γ can be calculated as

$$\gamma = \frac{C - D}{C + D} \tag{7}$$

where C is the total number of concordant pairs of observations and D is the total number of discordant pairs. The properties of gamma (and $\hat{\gamma}$) follow directly from its definition. Like the correlation, its range is $-1 < \gamma < 1$. The γ statistics were calculated for age-severity and speed ratio–severity associations. The $\hat{\gamma}$ values are 0.066 and 0.078 for age and speed ratio, respectively. Both statistics are significant at $\alpha = 0.05$ (p-value = 0.013 and 0.004, respectively).

For nominal/ordinal variables, a χ^2 test can be used to test the independence between the two variables. The χ^2 test results indicated that driver gender, fault, vehicle type, seat belt, point of impact, and area type are significant (p-value < 0.05), while alcohol usage, day of week, time, light condition, and weather condition are insignificant (p-value = 0.760, 0.250, 0.532, 0.377, and 0.556, respectively). It appears that drivers may take appropriate measures in bad weather conditions (or in bad light conditions). For example, they may decrease their speed and keep a safe distance from other vehicles. Finally, the insignificant variables were dropped from further investigation.

RESULTS

The number of cases used for the process of training is 2,336, of which 2,000 (85 percent) were used for the training phase and 336 (15 percent) for the testing phase. The ANN models are trained to map an input vector of dimensionality m (m factors) into an output vector of dimensionality 3 (i.e., a mapping problem from $\Re^m$ to $\Re^3$). The output patterns were coded as follows: (1 0 0) for no injury, (0 1 0) for possible injury/evident injury, and (0 0 1) for disabling injury/fatality. The input variables of the ANN have either numerical value or binary code. The coding for the input variables is described in Table 2. Finally, the performance of the network was measured in terms of network generalization and network size. Generalization performance of a network is defined as the percentage of patterns in the test set that are correctly classified by a trained network.

MLP Neural Networks

Before the training phase for the MLP was begun, all the input vectors were normalized to have a range of [−1, 1]. This vector normalization will help in fast convergence. The Neural Network Toolbox from the MATLAB library was used to train and test the

TABLE 1 Driver Injury Severity Distribution

Factor	No injury	Evident injury	Disabling injury	Total
Driver age				
15-24	288 (47.6%)	280 (46.4%)	36 (6.0%)	604
25-34	294 (50.6%)	254 (43.7%)	33 (5.7%)	581
35-44	231 (48.1%)	224 (46.7%)	25 (5.2%)	480
45-54	125 (40.6%)	159 (51.6%)	24 (7.8%)	308
55-64	72 (43.4%)	82 (49.4%)	12 (7.2%)	166
65-74	42 (35.6%)	67 (56.8%)	9 (7.6%)	118
75-84	28 (45.9%)	32 (52.5%)	1 (1.6%)	61
85+	8 (44.4%)	10 (55.6%)	0 (0.0%)	18
Driver gender				
Male	711 (54.4%)	534 (40.9%)	62 (4.7%)	1307
Female	377 (36.6%)	574 (55.8%)	78 (7.6%)	1029
Alcohol				
Not drinking	1040 (46.4%)	1066 (47.6%)	134 (6.0%)	2240
Drinking	48 (50.0 %)	42 (43.7%)	6 (6.3%)	96
Driver fault				
Not at fault	571 (41.1%)	721 (52.0%)	96 (6.9%)	1388
At fault	517 (54.5%)	387 (40.8%)	44 (4.7%)	948
Using seat belt				
In use	1024 (47.9%)	1003 (46.9%)	111 (5.2%)	2138
Not in use	64 (32.3%)	105 (53.0%)	29 (14.7%)	198
Vehicle type				
Passenger car	776 (43.0%)	902 (49.9%)	129 (7.1%)	1807
Passenger van	84 (50.3%)	79 (47.3%)	4 (2.4%)	167
Pickup/light truck	228 (63.0%)	127 (35.1%)	7 (1.9%)	362
Speed ratio				
< 0.25	156 (49.0%)	144 (45.3%)	18 (5.7%)	318
0.25-0.5	246 (49.8%)	221 (44.7%)	27 (5.5%)	494
0.5-0.75	199 (48.5%)	186 (45.4%)	25 (6.1%)	410
0.75-1	219 (47.1%)	217 (46.7%)	29 (6.2%)	465
1-1.25	254 (41.8%)	315 (51.8%)	39 (6.4%)	608
> 1.25	14 (34.1%)	25 (61.0%)	2 (4.9%)	41
Point of impact				
Not driver side	992 (50.0%)	891 (44.9%)	101 (5.1%)	1984
Driver side	96 (27.3%)	217 (61.6%)	39 (11.1%)	352
Day				
Weekday	800 (46.0%)	827 (47.6%)	112 (6.4%)	1739
Weekend	288 (48.2%)	281 (47.1%)	28 (4.7%)	597
Area type				
Rural	463 (43.8%)	532 (50.3%)	63 (5.9%)	1058
Urban	625 (48.9%)	576 (45.1%)	77 (6.0%)	1278
Time				
Off peak	857 (46.8%)	868 (47.5%)	104 (5.7%)	1829
Morning peak	91 (43.5%)	100 (47.9%)	18 (8.6%)	209
Evening peak	140 (47.0%)	140 (47.0%)	18 (6.0%)	298
Light				
Daylight	705 (46.0%)	737 (48.1%)	91 (5.9%)	1533
Dark (street light)	298 (50.0%)	264 (44.3%)	34 (5.7%)	596
Dark (no street light)	51 (41.8%)	61 (50.0%)	10 (8.2%)	122
Dusk/dawn	34 (40.0%)	46 (54.1%)	5 (5.9%)	85
Weather				
Clear	755 (46.0%)	782 (47.7%)	103 (6.3%)	1640
Not clear	333 (47.9%)	326 (46.8%)	37 (5.3%)	696

TABLE 2 Data Coding of the Input Variables

Factor	Binary/numerical code
Driver age	Numerical value
Driver gender	0 male; 1 female
Driver fault	0 driver not at fault; 1 driver at fault
Seat belt	0 in use ; 1 not in use
Vehicle type	(1, 0) Passenger car; (0, 1) Passenger van; (0, 0) pickup/light truck
Speed ratio	Numerical value
Point of impact	1 driver side; 0 not driver side
Area type	0 urban; 1 rural

MLP (*21*). The training algorithm was the Levenberg-Marquardt algorithm. This algorithm starts from an arbitrary set of interconnection weights randomly chosen from a uniform distribution of values between 0 and 1; the algorithm then tries to minimize the differences between the network output and the desired outputs. All runs had been carried out with a maximum number of epochs (a complete list presentation) of 100 and with 0.0001 *MSE* as the goal value. The input layer has nine neurons that represent eight different factors, and the output layer has three neurons that represent the three injury levels. All transfer functions at the hidden layer and the output layer are hyperbolic tangent sigmoid transfer functions. This type of transfer function has the following form:

$$g(\text{net}) = \frac{1 - \exp(-a\ \text{net})}{1 + \exp(-a\ \text{net})} \tag{8}$$

where net is the net input to a node and *a* is a positive slope parameter that gives different looks to the sigmoid function. To select the number of hidden nodes, an experiment with different values of hidden nodes (5 to 25 hidden nodes in increments of 5) was carried out. The number of hidden nodes that gives the best network performance was selected. Furthermore, for each value of hidden nodes, three different initial sets of interconnection weights were used to investigate the stability and the convergence speed of the network. The final model has 9 input nodes, 15 hidden nodes, and 3 output nodes. The generalization performance (classification accuracy) of the model is 65.6 and 60.4 percent for the training and testing phases, respectively.

Fuzzy ARTMAP Neural Network

Before training began, all training patterns were normalized and complement coded to have values in the range [0, 1]. These preprocessing operations are required by the fuzzy ARTMAP algorithm. Since the generalization performance of fuzzy ARTMAP depends on the order of pattern presentation in the training phase, 10 different random orders of pattern presentation were investigated. A visual C++ code was written to train and test the fuzzy ARTMAP ANN. The worst generalization performance was 40.5 percent and the best was 56.2 percent. The model that gives the best generalization performance has 272 nodes in the ART_a module and 3 nodes in the ART_b module.

To further improve the generalization performance, the O-ARTMAP was used. Before training began, all patterns went through an ordering algorithm, the purpose of which was to identify the order in which to present training patterns during the training phase of fuzzy ARTMAP. This task was accomplished by the K-means algorithm. The number of cluster centers was arbitrarily chosen to be 10. This order of training patterns was then presented to fuzzy ARTMAP to start the training phase. The O-ARTMAP gave a generalization performance of 58.1 percent. The size of the network that O-ARTMAP created was 285 nodes at the ART_a module and 3 nodes at the ART_b module. Therefore, the O-ARTMAP model will be considered as the final model for fuzzy ARTMAP neural networks.

MLP Versus O-ARTMAP

The MLP and the O-ARTMAP were compared to determine which ANN architecture better describes the data. Generally, the MLP and O-ARTMAP neural networks can be compared in terms of generalization performance and network size (*18*). The MLP gave better generalization performance (65.6 and 60.4 percent for training and testing, respectively) than the O-ARTMAP (58.1 percent). The MLP had a smaller network size; there were 15 hidden nodes in the MLP compared with 352 nodes in the ART_a module of the O-ARTMAP. Therefore, the MLP was selected as the ANN model that better describes the data. The performance of this MLP model was compared with that of an ordered logit model. The classification accuracy of the ordered logit model was 58.9 and 57.1 percent for the training and testing phases, respectively, indicating again that the MLP model has better performance.

To further check the performance of the selected MLP neural network model, it was tested with the 1996 accident data for the Central Florida area. These data include all two-vehicle accidents that were reported at signalized intersections. A total of 1,870 drivers were involved in 935 traffic accidents reported during 1996; 960 driver involvements (51.4 percent) resulted in no injury, 814 (43.5 percent) in possible/evident injury, and 96 (5.1 percent) in disabling injury. The classification accuracy of the MLP model with the 1996 data is 63.7 percent.

Injury Severity Model Simulations

To start prediction using the MLP neural network model, a simulation experiment with all possible combinations of the eight input factors was carried out. The use of this simulation experiment helps in understanding the ANN model. The simulated input patterns were created using all possible combinations of the binary input factors (area type, point of impact, driver fault, driver gender, seat belt usage), 3 types of vehicle, 16 levels of driver age (ages 15–90, in increments of 5 years), and 6 speed ratio categories (ranging from 0.25 to 1.50, with increments of 0.25). As a result, the experiment has 9,216 input patterns. This experiment extends the investigation of injury severity to cases not included in the training set. The task of the MLP model is to assign an output severity level for each input pattern, which enables an understanding as to which factors lead to severe injuries.

A sample of the simulated experiment is presented in Figure 3. This figure shows the MLP responses (i.e., predicted injury severity level) for a case of a driver not at fault, wearing a seat belt, in a passenger car, and with point of impact not at the driver's side. The figure depicts the relationship between driver age and gender, speed ratio, and the type of intersection (rural or urban). Speed ratio has a significant effect on injury severity; as the speed ratio increases the injury severity level increases. In addition, older drivers tend to have a greater risk of getting injured in traffic accidents than do younger drivers. Rural intersections are more dangerous than urban intersections in terms of driver injury severity. Figure 3 also shows that a female driver has a greater chance of experiencing a severe injury than does a male driver.

Similar graphs were constructed by changing one factor at a time. For example, Figure 3 can be drawn for the same conditions, but with the driver at fault, or not wearing a seat belt, or any combination of other factors. Results show that at-fault drivers have less likelihood of experiencing a severe injury than do those not at fault. The role of at-fault drivers in striking other vehicles makes them less likely to be injured than drivers of the other vehicles. Wearing a seat belt decreases the chance of having severe injuries. Vehicle type plays a role in driver injury severity. Drivers of passenger cars are more likely to experience higher injury severity than are those of passenger vans or pickup trucks. Finally, drivers exposed to impact at their side experience a higher injury severity than those exposed to impact elsewhere.

CONCLUSIONS

ANN models from the field of artificial intelligence have recently received increased attention in the field of transportation. Their adaptive nature and learning capabilities have been considered the most important features that make ANNs superior to other traditional techniques. This study presents a modeling technique that applies artificial ANNs to predict driver injury severity in traffic accidents at signalized intersections. Two well-known ANN paradigms were investigated: the MLP and fuzzy ARTMAP neural networks. In addition, the fuzzy ARTMAP was enhanced with an ordering algorithm

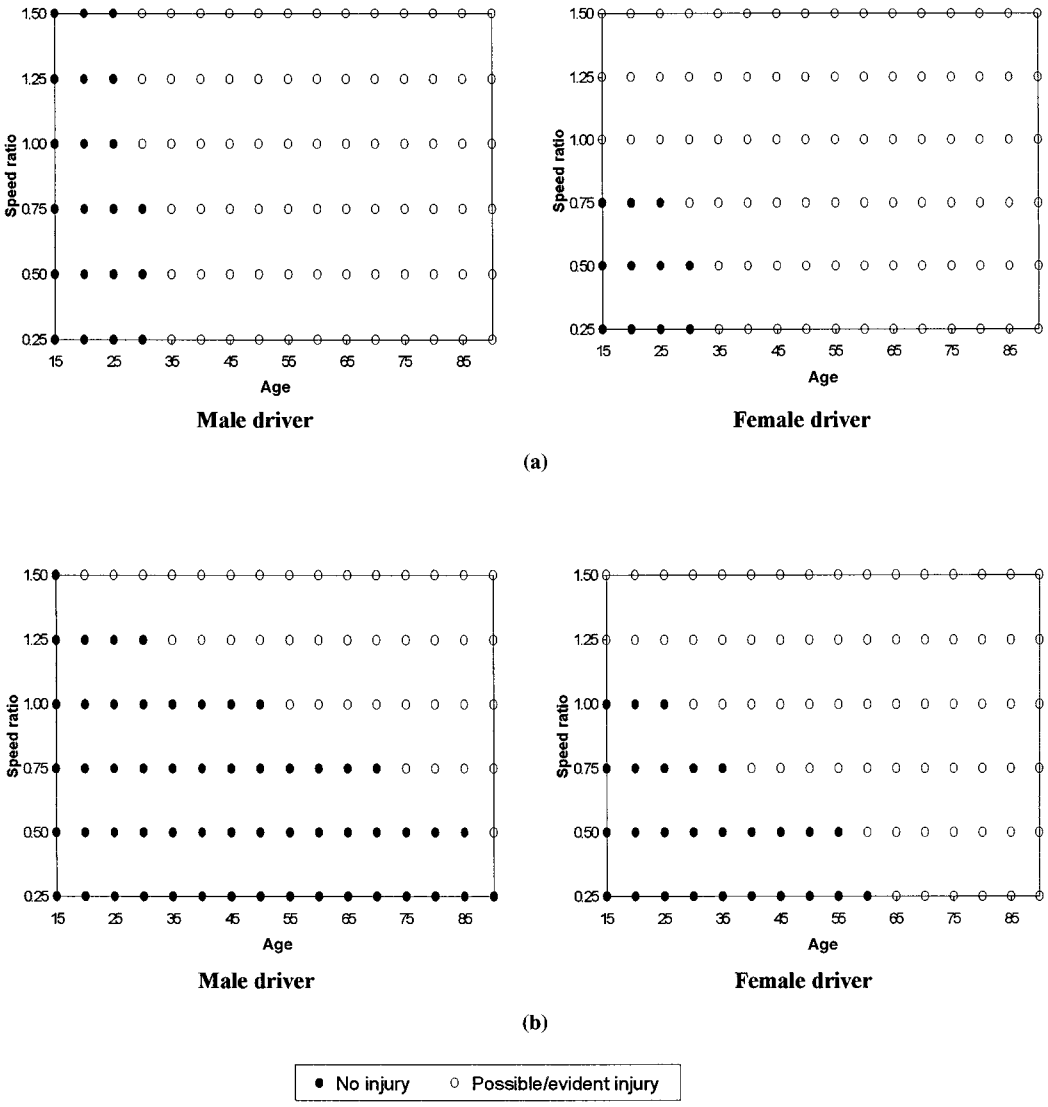

FIGURE 3 Distribution of predicted injury severity using the MLP neural network for (*a*) rural intersections and (*b*) urban intersections (with driver not at fault, wearing a seat belt, and driving a passenger car, and with point of impact not at driver's side).

that improves the network performance. The classification accuracy of the MLP model is 65.6 and 60.4 percent for the training and testing phases, respectively. The fuzzy ARTMAP neural network has a classification accuracy of 56.2 percent. The O-ARTMAP gave a classification accuracy of 58.1 percent. Therefore, the MLP model was selected as the best ANN model that fits the data. Furthermore, a calibrated ordered logit model confirms the results obtained from the MLP neural network. The ordered logit model correctly classified only 58.9 and 57.1 percent for the training and testing phases, respectively. Therefore, the MLP neural network provided the best training (calibration) and testing performance. This shows that MLP networks in particular, and ANNs in general, have promising potential in modeling injury severity.

To start prediction using the MLP neural network model, a simulation experiment with all possible combinations of the eight input factors was carried out. This simulation experiment helps provide an understanding of the ANN model. The experiment extends the investigation of injury severity to cases not included in the training set. The results show that rural intersections are more dangerous in terms of driver injury severity than urban intersections. In addition, female drivers are more likely to experience a severe injury than are male drivers. Speed ratio increases the likelihood of injury severity. At-fault drivers have less likelihood of experiencing severe injury than do those not at fault. Wearing a seat belt decreases the chance of having severe injuries. The vehicle type plays a role in driver injury severity. Drivers of passenger cars are more likely to experience a higher injury severity level than are drivers of passenger vans or pickup trucks. Finally, drivers exposed to impact at their side experience a higher injury severity level than those exposed to impact elsewhere.

For the data used in this analysis, the MLP neural network model provided better training and testing performance than other ANNs and statistical techniques. This shows that MLP in particular, and ANNs in general, have promising potential in modeling injury

severity. On the basis of the analysis presented in this paper, further investigation of the use of different ANN architectures (including MLP) in other transportation safety applications is suggested.

ACKNOWLEDGMENT

The authors acknowledge the comments and suggestions of three anonymous referees. Their recommendations resulted in a substantially improved paper.

REFERENCES

1. Lui, K. J., D. McGee, P. Rhodes, and D. Pollock. An Application of a Conditional Logistic Regression to Study the Effects of Safety Belts, Principal Impact Points, and Car Weights on Drivers' Fatalities. *Journal of Safety Research*, Vol. 19, No. 4, 1988, pp. 197–203.
2. Nassar, S., F. Saccomanno, and J. Shortreed. Road Accident Severity Analysis: A Micro Level Approach. *Canadian Journal of Civil Engineering*, Vol. 21, 1994, pp. 847–855.
3. Kim, K., L. Nitz, J. Richardson, and L. Li. Analyzing the Relationship Between Crash Types and Injuries in Motor Vehicle Collisions in Hawaii. In *Transportation Research Record 1467*, TRB, National Research Council, Washington, D.C., 1994, pp. 9–13.
4. Kim, K., L. Nitz, J. Richardson, and L. Li. Personal and Behavioral Predictors of Automobile Crash and Injury Severity. *Accident Analysis and Prevention*, Vol. 27, No. 4, 1995, pp. 469–481.
5. Shankar, V., F. Mannering, and W. Barfield. Statistical Analysis of Accident Severity on Rural Freeways. *Accident Analysis and Prevention*, Vol. 28, No. 3, 1996, pp. 391–401.
6. Edwards, J. The Relationship Between Road Accident Severity and Recorded Weather. *Journal of Safety Research*, Vol. 29, No. 4, 1998, pp. 249–262.
7. Kikuchi, S., R. Nanda, and V. Perincherry. A Method To Estimate Trip O-D Patterns Using a Neural Network Approach. *Transportation Planning and Technology*, Vol. 17, No. 5, 1993, pp. 51–65.
8. Dougherty, M. A Review of Neural Networks Applied to Transport. *Transportation Research C*, Vol. 3, No. 4, 1995, pp. 247–260.
9. Subba, R., P. Sikdar, K. Krishna, and S. Dhingra. Another Insight into Artificial Neural Networks Through Behavioral Analysis of Access Mode Choice. *Computers, Environment and Urban Systems*, Vol. 22, No. 5, 1998, pp. 485–496.
10. Mozolin, M., J. Thill, and E. Usery. Trip Distribution Forecasting with Multilayer Perceptron Neural Networks: A Critical Evaluation. *Transportation Research B*, Vol. 34, No. 4, 2000, pp. 53–73.
11. Mussone, L., A. Ferrari, and M. Oneta. An Analysis of Urban Collisions Using an Artificial Intelligence Model. *Accident Analysis and Prevention*, Vol. 31, 1999, pp. 705–718.
12. Lin, C., and G. Lee. *Neural Fuzzy Systems: A Neuro-Fuzzy Synergism to Intelligent Systems*. Prentice-Hall, Upper Saddle River, N.J., 1996.
13. Hashemi, R., L. Blanc, C. Rucks, and A. Shearry. Neural Network for Transportation Safety Modeling. *Expert Systems with Applications*, Vol. 9, 1995, pp. 247–256.
14. Nijkamp, P., A. Reggiani, and T. Tritapepe. Modeling Inter-Urban Transport Flows in Italy: A Comparison Between Neural Network Analysis and Logit Analysis. *Transportation Research C*, Vol. 4, No. 6, 1996, pp. 323–338.
15. Rumelhart, D. E., G. E. Hinton, and R. J. Williams. *Learning Internal Representation by Error Propagation, Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1: Foundations (D. E. Rumelhart and J. L. McClelland, eds.), MIT Press, Cambridge, Mass., 1986, pp. 318–362.
16. Carpenter, G. A., S. Grossberg, N. Markuzon, J. H. Reynolds, and D. B. Rosen. Fuzzy ARTMAP: A Neural-Network Architecture for Incremental Supervised Learning of Analog Multidimensional Maps. *IEEE Transactions on Neural Networks*, Vol. 3, 1992, pp. 698–713.
17. Grossberg, S. Adaptive Pattern Recognition and Universal Recording II: Feedback, Expectation, Olfaction, and Illusions. *Biological Cybernetics*, Vol. 23, 1976, pp. 187–202.
18. Dagher, I., M. Georgiopoulos, G. Helileman, and G. Bebis. An Ordering Algorithm for Pattern Presentation in Fuzzy ARTMAP that Tends to Improve Generalization Performance. *IEEE Transactions on Neural Networks*, Vol. 10, 1999, pp. 768–778.
19. Tou, J. T., and R. C. Gonzalez. *Pattern Recognition Principles*. Addison-Wesley, Reading, Mass., 1976.
20. Agresti, A. *Categorical Data Analysis*. John Wiley & Sons, New York, 1990.
21. Demuth, H., and M. Beale. *Neural Network Toolbox for Use with MATLAB: User's Guide Version 3.0*. The MathWorks, Inc., PWS Publishing, Boston, Mass., 1998.

Publication of this paper sponsored by Committee on Safety Data, Analysis, and Evaluation.