IMAN AGHAYAN, Ph.D. Candidate
E-mail: iman.aghayan@cc.emu.edu.tr
Eastern Mediterranean University
Dept. of Civil Engineering
Gazimagusa, Mersin 10, Turkey
NIMA NOII, M.Sc.
E-mail: nima.noii@myport.ac.uk
University of Portsmouth
School of Civil Engineering and Surveying
Portsmouth, Hampshire, United Kingdom
MEHMET METIN KUNT, Ph.D.
E-mail: metin.kunt@emu.edu.tr
Eastern Mediterranean University
Dept. of Civil Engineering
Gazimagusa, Mersin 10, Turkey

Science in Traffic and Transport
Original Scientific Paper
Accepted: Nov. 25, 2011
Approved: Nov. 13, 2012

EXTENDED TRAFFIC CRASH MODELLING THROUGH PRECISION AND RESPONSE TIME USING FUZZY CLUSTERING ALGORITHMS COMPARED WITH MULTI-LAYER PERCEPTRON

ABSTRACT

This paper compares two fuzzy clustering algorithms – fuzzy subtractive clustering and fuzzy C-means clustering – to a multi-layer perceptron neural network for their ability to predict the severity of crash injuries and to estimate the response time on the traffic crash data. Four clustering algorithms – hierarchical, K-means, subtractive clustering, and fuzzy C-means clustering – were used to obtain the optimum number of clusters based on the mean silhouette coefficient and R-value before applying the fuzzy clustering algorithms. The best-fit algorithms were selected according to two criteria: precision (root mean square, R-value, mean absolute errors, and sum of square error) and response time (t). The highest R-value was obtained for the multi-layer perceptron (0.89), demonstrating that the multi-layer perceptron had a high precision in traffic crash prediction among the prediction models, and that it was stable even in the presence of outliers and overlapping data. Meanwhile, in comparison with other prediction models, fuzzy subtractive clustering provided the lowest value for response time (0.284 second), 9.28 times faster than the time of multi-layer perceptron, meaning that it could lead to developing an on-line system for processing data from detectors and/or a real-time traffic database. The model can be extended through improvements based on additional data through induction procedure.

KEYWORDS

fuzzy subtractive, fuzzy C-means, hierarchical clustering, K-means clustering, multi-layer perceptron, traffic crash severity

1. INTRODUCTION

In recent years, a dramatic increase in traffic accidents worldwide has brought the problem of improving traffic safety to the attention of health officials who now approach the problem as they would a biological disease. Road traffic accidents 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 an accident and the influencing factors is nonlinear and complicated; it cannot be described with an explicit mathematical model.

One of the most important tools for investigating the relationship between crash occurrence and traffic risk factors is a crash prediction model. In this paper, seven variables are selected as input for such a model: driver's gender, driver's age, crash time, type of vehicle, weather conditions, trafficway characteristic, and collision type. The selected output variable is the injury severity, which consists of three levels: no injury, evident injury, and disabling injury/fatality. Two fuzzy clustering algorithms and a multi-layer perceptron are used to determine the suitability of those input variables and injury severity levels for model predictions.

The first prediction model is an artificial neural network (ANN). A neural network is composed of simple elements operating in parallel, as found in biological nervous systems. As in nature, the connections between the elements largely determine the network function. In this study, a multi-layer perceptron (MLP) neural network architecture that consists of a multi-layer feed-forward network with sigmoid hidden neurons and linear output neurons is used.

Clustering techniques focus on obtaining useful information by the grouping of multi-dimensional data into clusters. In this study, four clustering algorithms – hierarchical, K-means, subtractive clustering, and fuzzy C-means (FCM) clustering – are used to obtain the optimum number of clusters.

The second prediction model for determining the suitability of the input variables and output variables is a fuzzy inference system (FIS) using FCM clustering based on Takagi-Sugeno-Kang (TSK) and Mamdani. The FCM clustering model optimizes the objective function to obtain the membership degree for each sample point relative to all the cluster centres, then determines the generic of the sample points, and finally achieves automatic classification for data samples. The third prediction model is a FIS using subtractive clustering based on the TSK-FIS structure. The aim of subtractive clustering is to estimate both the number and initial locations of cluster centres and extract the TSK fuzzy rules from the input/output variables.

The first objective of this study is to obtain the optimum number of clusters based on the clustering algorithms before conducting an analysis with the fuzzy clustering algorithms. The second objective of this study is to determine the most suitable prediction model from among the tested models based on two criteria: precision (root mean square (RMSE), R-value, mean absolute errors (MAE), and sum of square error (SSE)) and response time (t). The precision factor establishes if a model is able to accurately predict traffic crash severity, while the response time establishes if the model can produce results in a reasonable period of time. Thus, prediction models are chosen based on having the highest precision and the lowest response time. The third objective of this study is to create a model that can be updated with additional data beyond whatever is previously used so that the prediction models are improved based on new information through induction procedure.

2. RELATED WORK

The common models used in traffic safety are the traditional Poisson and Poisson-gamma models. Those models are applied for modelling discrete, independent, and non-negative events. The Poisson and Poisson-gamma models have been used for predicting motor vehicle crashes [1, 2]. Other statistical models applied to accident data include the binomial, the zero-inflated Poisson (ZIP), the zero-inflated negative binomial (ZINB), and the multinomial probability models.

ANN has been verified to be efficient in many other fields, but have seldom been used as a modelling approach in the analysis of crash-related injury severity. In the transportation field, Mussone et al. [3] applied ANN to analyze vehicular crashes at an intersection in Milan, Italy. A number of studies have used ANN to recognize groups of drivers at greater risk of being injured or killed in traffic crashes [4, 5]. Other applications of ANN have included traffic prediction [6, 7], estimation of traffic parameters [8], incident detection [9, 10], travel behaviour analysis [11-13], and traffic accident analysis [14, 3, 15, 16]. Abdelwahab and Abdel-Aty [17] used ANN to provide the association of driver injury severity and crash factors such as the driver, vehicle, roadway, and environmental characteristics. A probabilistic neural network (PNN) model was applied by Abdel-Aty and Pande [16] for crash prediction on the Interstate 4 corridor in Orlando, Florida. Kunt et al. [18] used an ANN, a genetic algorithm, and a genetic algorithm combined with pattern search for predicting the severity of freeway traffic crashes.

Zadeh [19] introduced fuzzy logic in the 1960s. There are a series of justifications for using fuzzy logic in the modelling of complex processes. Fuzzy set theory techniques have been used in crash prevention efforts. Akiyama and Sho [20] studied the traffic safety problem on urban expressways. Hadji Hosseinlou and Aghayan [21] used fuzzy logic to predict the traffic crash severity on the Tehran-Ghom freeway in Iran. Fuzzy logic has been used for the control of traffic systems [22-25]. The combination of fuzzy logic and neural networks has been applied for incident detection on freeways by Ishak and Al-Deek [26]. Ruspini [27] was the first to propose fuzzy c-partitions as a fuzzy approach for clustering, and then, the FCM algorithms were modified by Dunn [28] and generalized by Bezdek [29]. In connection with FCM algorithm, Sugeno and Yasukawa [30] determined the optimal number of clusters in the output space. Chen et al. [31] suggested the data space should be classified with regard to the input data in addition to linear relationships between input and output data. Feature-weighted FCM based on feature selection methods and on competitive agglomeration were proposed by Wang et al. [32] and Frigui and Nasraoui [33], respectively.

Chiu [34] introduced subtractive clustering. In subtractive clustering, data points were selected for cluster centres to solve computational difficulties that can arise in mountain clustering when problem dimensions are suitably increased for handling large data sets. Yager and Filev developed the mountain method for estimating cluster centroids (Yager and Filev, [35]). Hayajneh and Hassan [36] applied a fuzzy subtractive (FS) clustering and a FIS based on the Sugeno method for drilling processes.

One of the clustering methods is the K-means algorithm [37]. K-means clustering is an unsupervised pattern classification method. Fukunaga [38] used K-means clustering on continuous data. Pena et al. [39] applied various methods for process of initializing in the K-means algorithm. The K-means algorithm performance is related to the initial cluster centres; thus, Khan and Ahmad [40] and Redmond and Heneghan [41] suggested an algorithm for K-means clustering to determine initial cluster centres. Hierarchical clustering according to agglomerative algorithms has been developed [42-46], where objects are initially allocated to their own cluster in addition to pairs of clusters that are merged repeatedly until a tree is completely formed.

3. METHODOLOGY

3.1 Parameters for prediction

The dataset used in this study consists of 1,049 traffic crashes and was derived from traffic crashes reported between 2005 and 2010 on the North Cyprus primary road network. The dataset includes only crash data that are complete with regard to all input variables that were used in this study. These data were used as training and checking data for the MLP, FCM clustering, and FS clustering as well as a comparison for the predictions from all three models. Three injury levels were taken into the consideration for this study: no injury, evident injury, disabling injury/fatality, and seven input variables were selected from the data. *Table 1* shows the input and output variables. The performances of the three modelling approaches (MLP, FCM clustering, and FS clustering) were obtained using MATLAB software.

Table 1 - Description of the Study Variables.

Input Variable		Coding/Values	Data
1	Driver's Gender	Man Woman	82.28% 17.72%
2	Driver's Age	Year	-
3	Crash Time	Day Night	67.17% 32.83%
4	Type of Vehicle	Passenger car Pick-up	59.76% 40.24%
5	Weather Condition	Clear Cloudy Rainy	95.19% 1.81% 3.00%
6	Trafficway Character	Curve Straight road segment	30.73% 69.27%
7	Collision Type	Rear-end Right-angle Side-wipe	12.81% 25.42% 61.77%
Output Variable			
1	Driver Injury Severity	No injury : {1,0,0} Evident injury : {0,1,0} Fatality : {0,0,1}	37.84% 59.75% 2.41%

3.2 Typical steps in designing the model

In this study, a comparison of MLP with FCM clustering and FS clustering was performed by considering

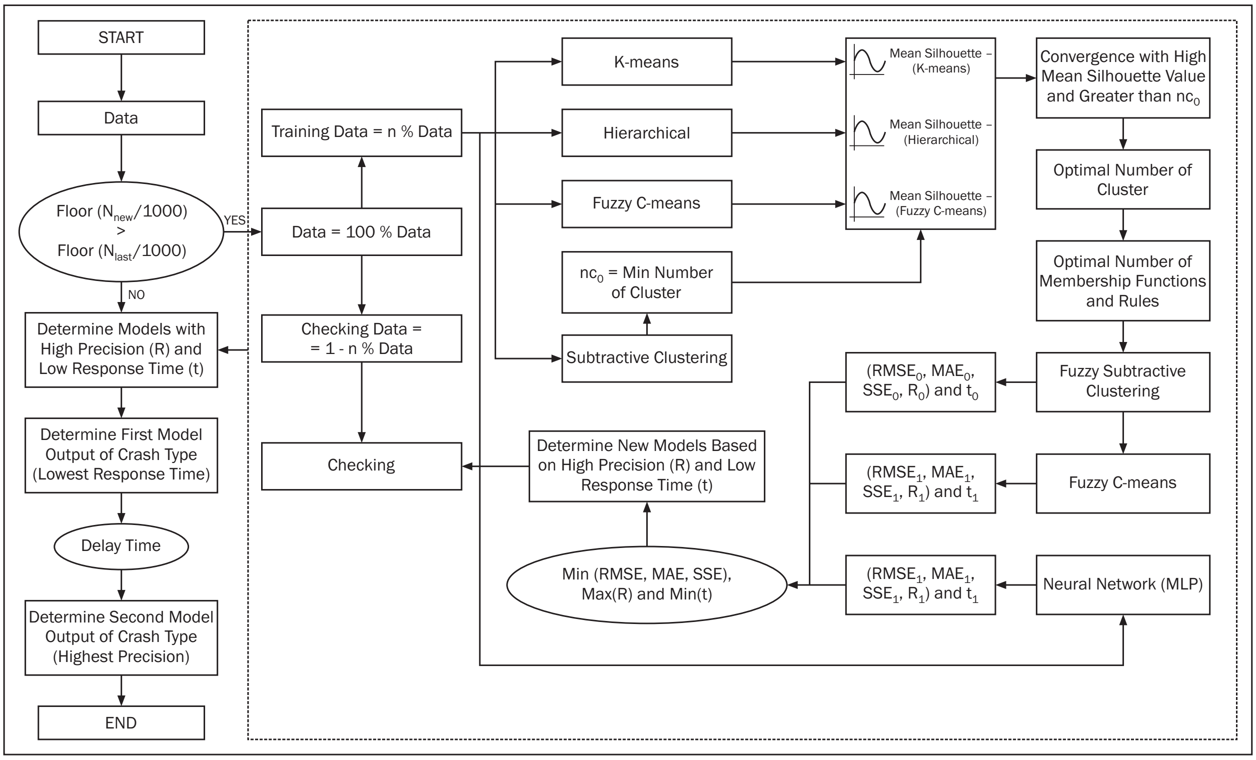

Figure 1 - Flowchart for the processes in a typical run

the optimum number of data cluster algorithms employed for improving the traffic crash prediction procedure. The first modelling step was the training phase that used 70% of the data, and the other 30% of the data were used for model validation and testing to improve the model. The 1,049 records collected from police records were used to construct the initial prediction model. However, before initiating the main part of the flowchart shown in *Figure 1* by the dashed line, the number of available data records were checked because the model can be updated with every batch of 1,000 records. In other words, the model can be updated every additional 1,000 records beyond the preliminary data. This means the model has this ability to improve itself with new data. Hierarchical, K-means, subtractive clustering and FCM clustering were employed for obtaining the optimum number of clusters based on mean silhouette coefficient and R-value. Consequentially, the optimum number of clusters achieved before was used in FS and FCM. In addition, the best-fit prediction algorithms between the fuzzy clustering algorithms and MLP were selected based on two criteria: precision (R, RMSE, MAE, and SSE) and response time (t). This procedure led to identification of suitable models with respect to both response time (t) and precision (R). Thus, if a fast prediction model was the goal, then this procedure identified the prediction model with the lowest response time, but if precision was the concern, then the procedure found the prediction model with the highest precision based on checking the data before the model started performing predictions. The first model output had the lowest response time, while the second model output, delayed by a few seconds, had the highest precision.

3.3 Multi-layer perceptron neural network

This study used a multi-layer perceptron (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 2* 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 data that were randomly divided into sets for model training, testing, and validating. The MLP model had 7 inputs, 20 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 analyzed as a one-layer network with 20 inputs, 3 neurons, and a 3*20 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 error (LMS) 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*.

3.4 Hierarchical clustering

The agglomerative hierarchical algorithm was employed in this study. The agglomerative algorithm is initiated by assuming that each of *n* objects to be clus-

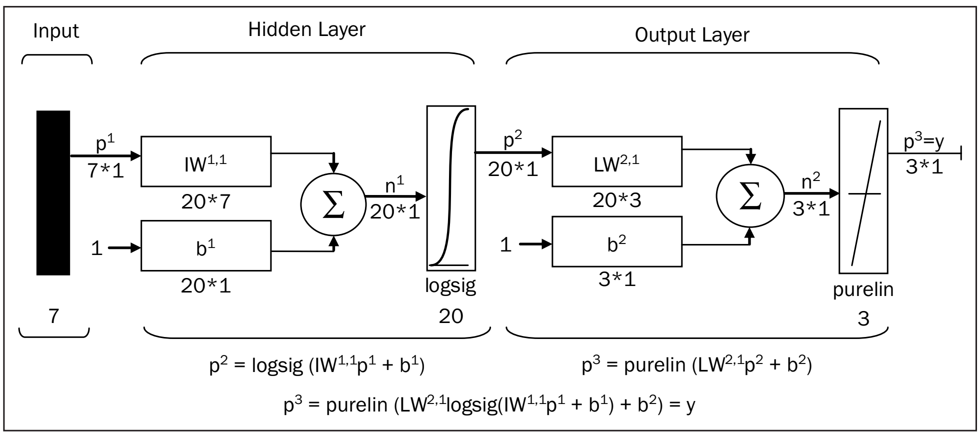

Figure 2 - Structure of final multi-layer-perceptron neural network model

tered is a unique cluster. The objects were compared with each other using a Euclidean distance to determine the distance between objects. That process was repeated until the number of clusters was obtained. The average linkage method defined in Eq. 1 was applied for comparing the clusters in each stage between all pairs of objects and deciding which of them should be combined.

$$d(r,s) = \frac{1}{n_r n_s}\sum_{i=1}^{n_r}\sum_{j=1}^{n_s}\| x_{ri} - x_{sj} \| \quad (1)$$

Here, x_{ri} is the i^{th} object in cluster r. This methodology partitions data by identifying natural groupings in the hierarchical tree or by cutting off the hierarchical tree at a random point. The cophenetic correlation coefficient (CPCC) for a cluster tree is defined as the linear correlation coefficient between the cophenetic distances obtained from the tree and the original distances (or dissimilarities), which varies between 0 and +1. The CPCC between Z, the average linkage method, and Y, the Euclidean distance for all data, is defined by Eq. 2:

$$c = \left| \frac{\sum_{i<j}(Y_{ij}-y)(Z_{ij}-z)}{\sum_{i<j}(Y_{ij}-y)^2 \sum_{i<j}(Z_{ij}-z)^2} \right| \quad (2)$$

where Y_{ij} is the distance(Y_i, Y_j), Z_{ij} is the cophenetic distance between objects i and j in Z and y and z are the average of Y and Z, respectively. In this study, the CPCC obtained for the preliminary data was 0.8425, which indicated that the hierarchical cluster tree was fairly good in terms of accuracy of the clustering solution.

3.5 K-means clustering

The K-means methodology depends on an initial vector to lead to an optimum solution. In this study, a modified K-means methodology was employed to reach the local minimum in any circumstance, which was useful for the large number of records. The modified K-means method involved a batch phase and online updates in which the first step entailed reassigning the vector in the closest cluster centroid through recalculation of cluster centroids and the second step entailed determining a clustering solution by convergence to a local minimum where points were individually reallocated and cluster centres were recalculated after each reallocation.

The pseudocode for K-means clustering is given in Algorithm 1.

3.6 Fuzzy C-means clustering

Similar to fuzzy rules, fuzzy clusters are well suited as a means for building a classification model. Clus-

Algorithm 1 - Modified K-means Clustering Algorithm

Clustering variables

X: An object; S_i: The i^{th} cluster; c_i: The centroid of cluster S_i; C: The centroid of all points;
N: The number of object in the data set; K: The number of clusters.

input: $X = \{x_1; x_2; \ldots; x_N\} \in \mathbb{R}^{N\times D}$ (N×D input data set)
output: $C = \{c_1; c_2; \ldots; c_K\} \in \mathbb{R}^{K\times D}$ (K cluster centres)
% replicates: Number of times to repeat the clustering, with a new set of initial cluster centroid
for (*replicates =1:1:rep*);
    Choose a random subset C of X as the initial set of cluster centres;
    while termination criterion is not met: $\{minimize: J(C) = \sum_{k=1}^{K}\sum_{j=1}^{c_i}\| x_i - c_k \|^2\}$
        for (*j=1:1:N*);
            Assign x_j to the nearest cluster;
            for (*i=1:1:K*);
                $S_i^t = \{x_j : \min(\| x_j - c_i^t \|, \| x_j - c_{i^*}^t \| \text{ for } i^* \in [1:K]-[i])\}$
            end
        end
        Recalculate the cluster centres;
        for (*k=1:1:K*)
            Cluster S_k^t includes the set of points x_i that are nearest to the centre c_k^{t+1};
            $|S_k^t| = \{x_i | S_k^t\}$; the number of data in cluster *i*;
            Calculate the new centre ck as the mean of the points that belong to S_k^t;
            $c_k^{t+1} = \frac{1}{|S_k^t|}\sum\{x_i \in S_k^t\}$
        end
    end
end
Best replicates: min {total sum of distances: [*1:rep*] }

ters are often considered as fuzzy rules to initialize a fuzzy rule system that is then optimized. The essential procedure of FCM is to find clusters such that the overall distance from a cluster prototype to each datum is minimized. The FCM algorithm is defined by the objective function:

$$J_{FCM}(U,V;X) = \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}^{m} \| x_k - v_i \|^2 \quad (3)$$

where $d_{ik}^2 = \| x_k - v_i \|^2$, and $\| x_k - v_i \|$ is the Euclidean distance between the centroids that characterizes the k^{th} data point and i^{th} cluster. Moreover, n is the number of data points, c the number of cluster, x_k is the k^{th} data point, v_i is the i^{th} cluster centre, and u_{ik} is the degree of membership of the k^{th} data point in the i^{th} cluster. The fuzziness parameter has a typical value of $m = 2$ [47]. The cluster centre v_i and the degree of membership function u_{ik} that are used in $J_{FCM}(U,V;X)$ are defined by:

$$u_{ik} = \frac{1}{\sum_{j=1}^{c}\left(\frac{\| x_k - v_i \|}{\| x_k - v_j \|}\right)^{\frac{2}{(m-1)}}} \quad (4)$$

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^{m} x_k}{\sum_{k=1}^{n} u_{ik}^{m}} \quad (5)$$

In this study, a modified FCM clustering was employed. For this procedure, the initial FCM partition was defined and set with the number of clusters equal to 3, the exponent for the partition matrix equal to 2, the maximum number of iterations equal to100 and minimum improvement equal to 1e-10. Based on this, the initial fuzzy cluster centres were calculated through the generation of the initial fuzzy partition. To improve the FCM clustering, the cluster centres and the membership grade points were updated, and the objective function defined in Eq. 3 was minimized to find the best location for each cluster. This procedure was terminated when the maximum number of iterations or minimum amount of improvement were reached.

3.7 Fuzzy subtractive clustering

Subtractive clustering uses data points as the candidates for cluster centres instead of grid points, which means that the computation is related to the problem size [48]. In fact, the cluster centres should be located at the data points to reduce the computation effort. The pseudocode for subtractive clustering is given in Algorithm 2. The pseudocode for FS clustering is given in Algorithm 3.

4. EXPERIMENTS AND EVALUATIONS

4.1 Comparison of clustering models

Use of a clustering methodology results in the optimum number of membership functions. *Figure 3* depicts the influence of the number of clusters along with the various radii in subtractive clustering. *Figure 4* shows the R-value for a given radius in the FS clustering algorithm. In Figures 3 and 4, the minimum number of clusters was 10, and the R-value achieved by the 10 clusters was 0.855. *Figure 5* shows the relationships between the number of clusters and the mean silhouette coefficient. It was found that, when the number of clusters was increased, the mean silhouette coefficient, which represents the overall quality of the clustering measurement, was decreased. *Figure 5* shows that the mean silhouette coefficients for hierarchical, K-means, and FCM clustering converged to 12 clusters. As explained above, by increasing the number of clusters, the R-value increased and the mean silhouette coefficient was decreased. Therefore, to satisfy two different evaluations for the cluster validity, 12 clusters were selected, which was more than the minimum number of 10 clusters obtained from subtractive clustering in *Figure 3*.

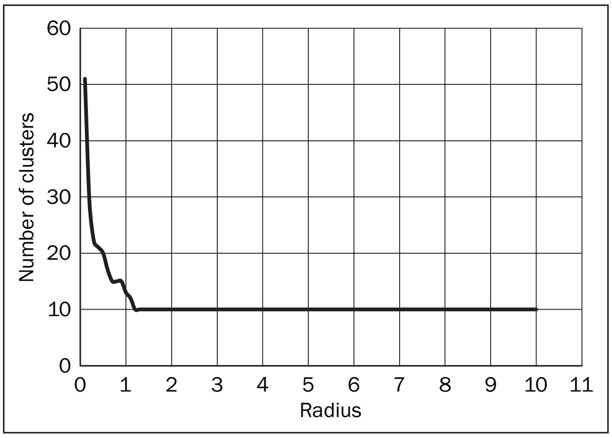

Figure 3 - The influence of the number of clusters with given radius in subtractive clustering

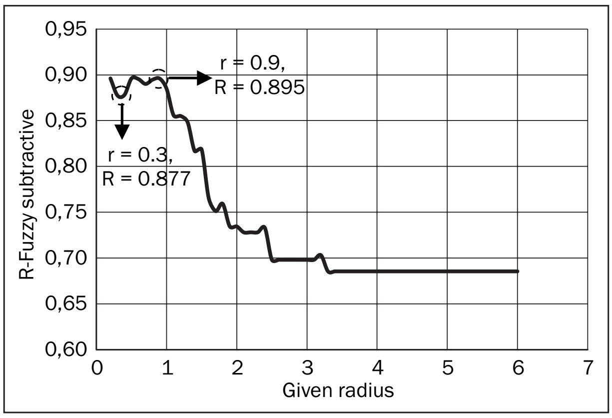

Figure 4 - R-values for given radii in FS

Algorithm 2 - Subtractive Clustering Algorithm

Clustering variables

X: An objects $(N \times D)$; x_i^*: location of i^{th} cluster; P_i^*: potential of i^{th} cluster; C: The number of clusters, r_j: radius of the j^{th} variable, r_a & r_b: positive constant, ε^{upper} = accept ratio, ε^{lower} = reject ratio.

input: $X = \{x_1; x_2; \ldots; x_N\} \in \mathbb{R}^{N \times D}$ $(N \times D$ input data set)

output: $X^* = \{x_1^*; x_2^*; \ldots; x_c^*\} \in \mathbb{R}^{C \times D}$ (c cluster centres)

for (*i,l=1:1:N*);

$$\| x_i - x_1 \| = \left(\sum_{j=1}^{D} \frac{(x_{ij} - x_{lj})^2}{r_j} \right)^{0.5}; \text{ \% scaled distance}$$

end

for (*i =1:1:N*);

$$P_1(x_i) = \sum_{j=1}^{N} \exp\left(-\alpha \| x_i - x_j \|^2\right) \text{ and } \alpha = \frac{4}{r_a^2}; \text{ \% initial potential for each data point}$$

end

$P_1^* = \text{argmax } x_{i=1}^N (P_1(x_i))$ % potential value for the first cluster centre.

$P_1^* = P_1(x_1^*)$ % location of the first cluster centre

while $(k \leq c); k = 2;$

$$P_k(x_i) = P_{k-1}(x_i) - P_{k-1}^* \exp\left(-\beta \| x_{k-1}^* - x_i \|^2\right) \text{ and } \beta = \frac{4}{r_b^2} \text{ \% next cluster center}$$

if $\varepsilon^{lower} P_1^* < P_k^* < \varepsilon^{upper} P_1^*$

$d_{\min} = \min(\| x_k^* - x_{1:k-1}^* \|)$

if $d_{\min}/r_a + P_k^*/P_1^* \geq 1$

$P_k^* = P_k(x_k^*)$;% location of the next cluster centre

continue; $k = k + 1$;% go to the beginning of the loop

else

$P_{k-1}(x_i \mid P_{k-1}^* = P_{k-1}(x_i)) = 0$;% eliminating rejected value by assigning potential value 0.

$P_{k-1}^* = \text{argmax } x_{i=1}^N (P_{k-1}(x_i))$; % choose next higher potential value

$P_{k-1}^* = P_{k-1}(x_{k-1}^*)$; return;

end

end

if $P_k^* > \varepsilon^{upper} P_1^*$

$P_k^* = P_k(x_k^*)$% location of the next cluster centre

continue; $k = k + 1$;%go to the beginning of the loop

elseif $P_k^* < \varepsilon^{lower} P_1^*$

Break; %the algorithm is finished.

end

end

Algorithm 3 - Fuzzy Subtractive Clustering-Rule i^{th}

Clustering variables

X: An objects $(N \times D)$; X^*: location of data cluster; Y: Input; Y^*: location of input cluster; Z: Output; Z^*: location of output cluster; A_j^i : Gaussian Membership function; B_j^i : Singleton Membership function

$X = \{Y, Z\} = \{x_1; x_2; \ldots; x_N\}; Y = \{y_1; y_2; \ldots; y_7\}; Z = \{z_1; z_2; z_3\}$

$X^* = \{Y^*, Z^*\} = \{x_1^*; x_2^*; \ldots; x_N^*\}; Y = \{y_1^*; y_2^*; \ldots; y_7^*\}; Z = \{z_1^*; z_2^*; z_3^*\}$

Rule$_i$: if $\{y_1 \text{ is } A_1^i \text{ and } y_2 \text{ is } A_2^i, \ldots, y_7 \text{ is } A_7^i\}$ then $\{z_1 \text{ is } B_1^i, z_2 \text{ is } B_2^i, z_3 \text{ is } B_3^i\}$, where:

$A_j^i(y_j) = \exp\left[\frac{-1}{2}\left(\frac{(y_j - y_{i,j}^*)}{\sigma_j^i}\right)^2\right]$ and $B_j^i(z_j) = \{1, \text{if: } y_j = y_{i,j}^* \text{ or } 0, \text{if: } y_j \neq y_{i,j}^*\}$, therefore:

$$\hat{Z} = \frac{\sum_{i=1}^{c} (\mu_i z_i^*)}{\sum_{i=1}^{c} (\mu_i)} = \frac{\sum_{i=1}^{c} \left(\exp\left(\alpha \| y - y_i^* \|^2\right) z_i^*\right)}{\sum_{i=1}^{c} \left(\exp\left(\alpha \| y - y_i^* \|^2\right)\right)} = \frac{\sum_{i=1}^{c} \left(\prod_{j=1}^{7} (A_j^i(y_j)) z_i^*\right)}{\sum_{i=1}^{c} \left(\prod_{j=1}^{7} (A_j^i(y_j))\right)}; \text{ \% Output vector}$$

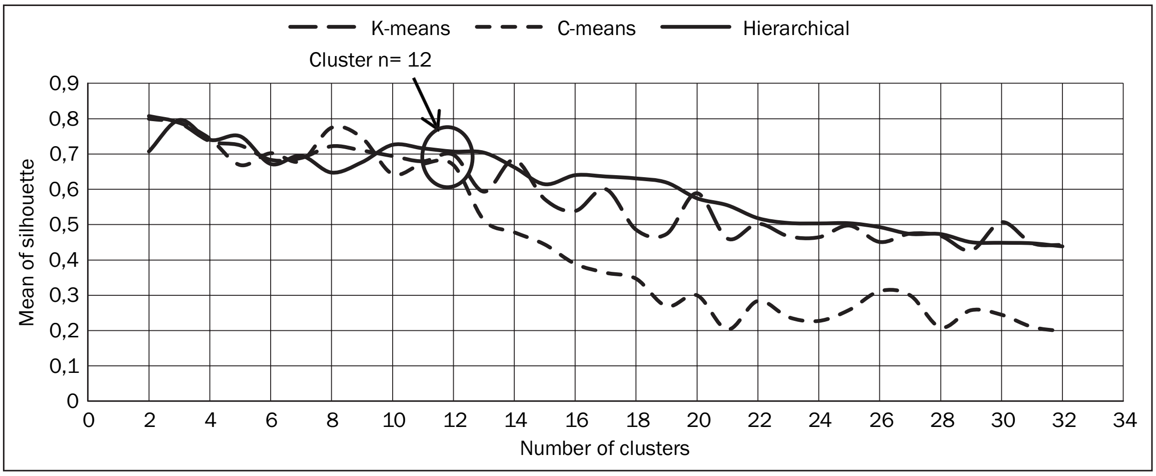

Figure 5 - Comparing the mean silhouette values in K-means, hierarchical, and FCM clustering

All clustering algorithms identified 12 clusters, which meant that each input and output was characterized by 12 membership functions. Moreover, the number of rules equals the number of clusters, and hence, 12 rules were created.

4.2 Fuzzy C-means clustering

Figure 6 shows an example of the membership functions (MFs) of the collision type and driver's age obtained from the FCM clustering algorithm. The actual and predicted values for 15% of the checking data are shown in *Figure 7*. The output values of the three levels (no injury: 1, evident injury: 2, fatality: 3) were either 0 or 1, as shown in *Table 1*. The mean response time of the FCM approach for 20 runs was 0.4744 seconds.

4.3 Fuzzy subtractive clustering

Subtractive clustering is a fast, one-pass algorithm to determine the approximate number of clusters and the cluster centres in the training dataset. However, in FS clustering, both input and output training data generates a Sugeno-type FIS structure.

In this case, the cluster radius indicates the range of influence of a cluster when the data space as a unit hypercube is considered. Specifying a small cluster radius usually yields many small clusters in the data and

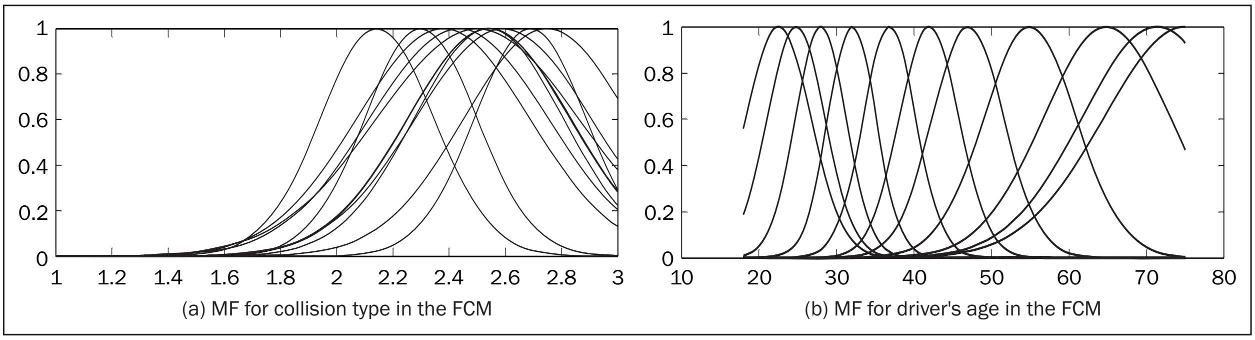

Figure 6 - Membership functions for input variables in FCM clustering.

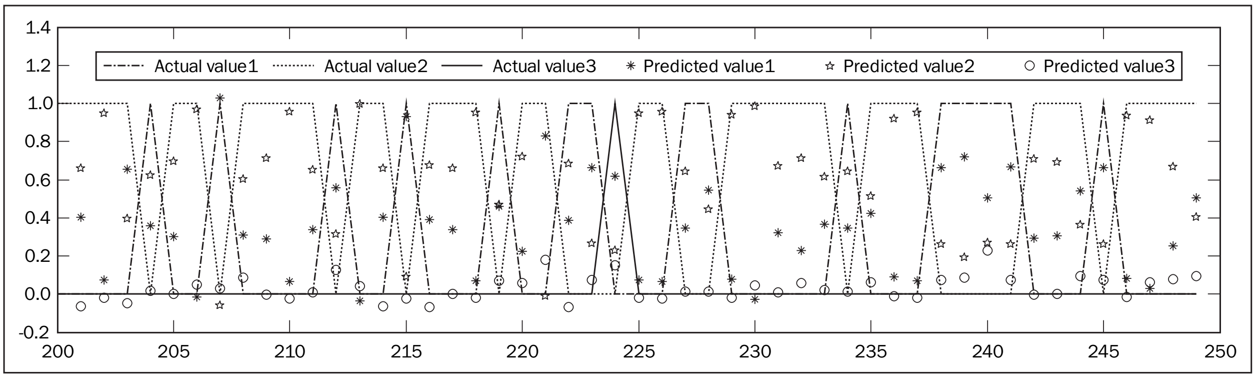

Figure 7 - Comparison of the actual and predicted values for checking data in FCM clustering

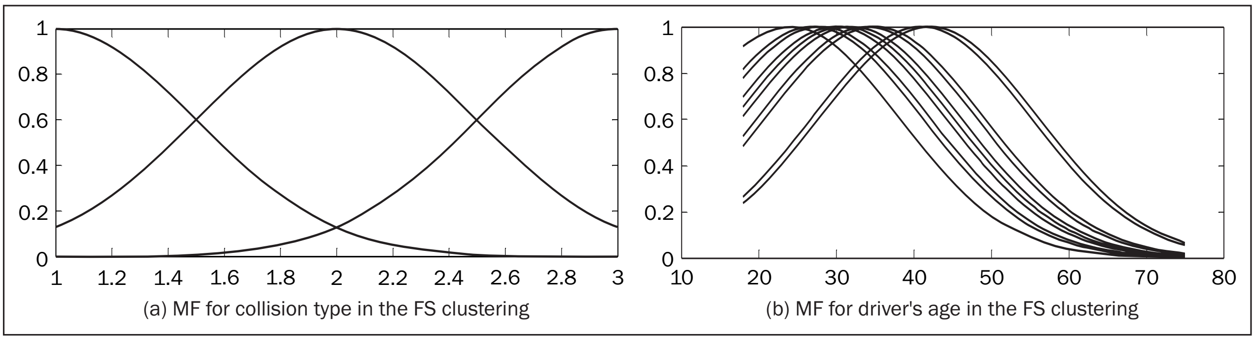

Figure 8 - Membership functions for input variables in FS clustering

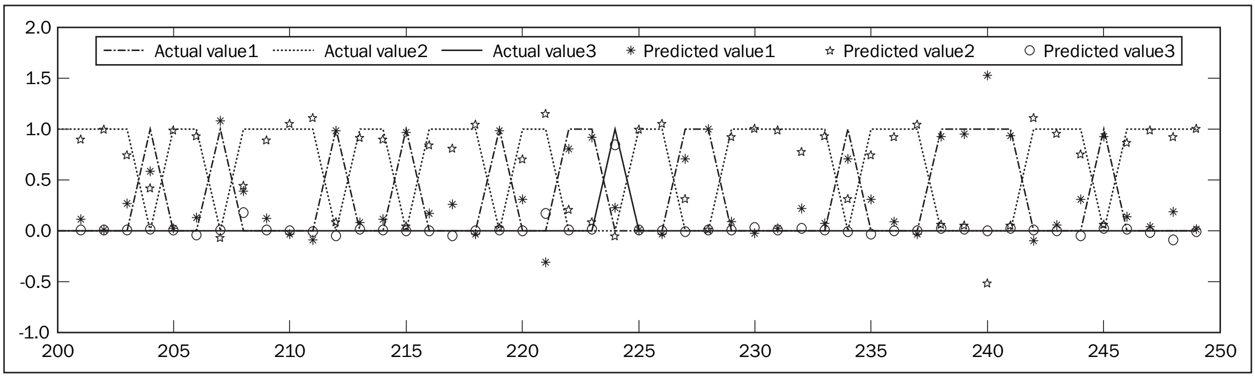

Figure 9 - Comparison of actual and predicted value for checking data in FS clustering

results in many rules. Specifying a large cluster radius usually yields a few large clusters in the data and results in fewer rules. *Figure 8* presents an example of the membership functions of the collision type and the driver's age as used in FS clustering. When compared with *Figure 6*, the figure reveals that the FS clustering algorithm had a lower computational cost than the FCM clustering because of a smaller number of membership functions and rules. The actual and predicted values based on the coding variables given in *Table 1* are shown in *Figure 9* for 15% of the checking data. In few circumstances the predicted values were out of the boundary condition defined to be between 0 and 1; meanwhile, overall similarity between actual and predicted values was not affected by errors happened in predicted values. The mean response time of the FS clustering for 20 runs was 0.2843 seconds.

4.4 MLP model results

The MLP, which was applied for training, testing, and validation, consisted of 7 inputs, 20 neurons in the hidden layers, and 3 neurons in the output layer. The data for training, validation, and testing of MLP application represented 70%, 15%, and 15% of all crash data, respectively. The results of the MLP model are shown in *Table 2* for 20 runs, which tabulates the prediction levels of injury severity patterns in the training, testing, and validation phases.

5. DISCUSSION

The relationship between the R-value and the number of clusters for FCM clustering calculated by the Mamdani and Sugeno fuzzy algorithms is represented in *Figure 10*. With the Sugeno fuzzy algorithm, the R-value was not related to the cluster number; whereas with the Mamdani fuzzy algorithm the R-value increased approximately with the number of clusters. *Figure 11* shows the relationship between the R-value and the number of clusters in the FS clustering. The R-value was 0.855 for the FS clustering with 12 clusters. Figures 10 and 11 show that increasing the number of clusters typically caused the R-value to increase, but

Table 2 - Prediction table for MLP model

R	No Injury	Evident Injury	Fatality	Overall
Training	0.7383	0.8819	0.8805	0.9102
Validation	0.6449	0.7208	0.7408	0.8115
Testing	0.5291	0.8259	0.8723	0.8547
All	0.6783	0.8623	0.8673	0.8920

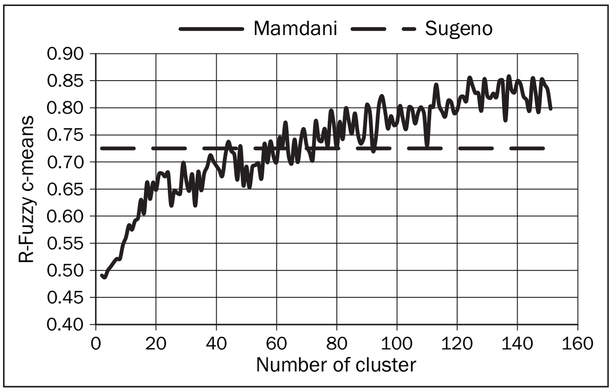

Figure 10 - R-values in FCM

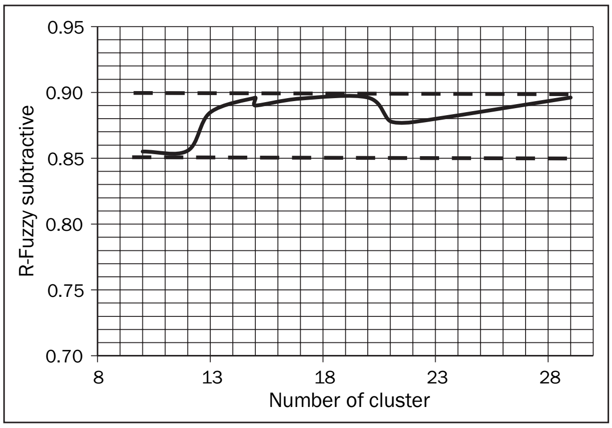

Figure 11 - R-values in FS

Figure 5 shows that such an increase in the number of clusters caused the mean silhouette coefficient to decrease. Subsequently, 10 was the optimum number of clusters determined by subtractive clustering as shown in *Figure 3*, the mean silhouette coefficient converged at 12 clusters, as shown in *Figure 5*. Then, because 12 is greater than 10, 12 clusters were selected for predicting the model.

Figure 12 represents the relationship between the number of program runs and the response time for the MLP, FS clustering and FCM clustering. The mean response time for FS clustering was less than that for the other models.

The MLP, FCM clustering, and FS clustering were compared on the basis of their R, MAE, SSE, and RMSE values, as shown in *Table 3*. The FS clustering provided the lowest elapsed time, 0.284 seconds, followed by FCM clustering, with an elapsed time of 0.474 seconds. FS clustering used the least amount of time with the precision that was less than MLP's precision, while MLP model used an exhaustive search in the greatest amount of time with the best precision. Thus, if a fast prediction model is the goal, FS clustering can be the right choice, but if precision is the main concern, then MLP is the best choice.

Figures 13, 14, and 15 display the residuals, which are the differences between the real values and predicted values for each model. In these figures, the abscissa shows 1,049 data points while the ordinate shows residuals for each cluster.

6. CONCLUSION

1. This study compared the FS clustering, FCM clustering, and MLP models to identify the model best suited for predicting traffic crash severity at three levels, fatality, evidence of injury, and no injury, as well as estimating the response time for processing traffic crash data.

Table 3 - Final results for the objective function in each model

Used Model	MAE	SSE	MSE	RMSE	R	T(sec)
Multi-Layer Perceptron (MLP)	0.129	132.927	0.044	0.211	0.892	2.635
Fuzzy Subtractive Clustering	0.148	159.62	0.060	0.245	0.855	0.284
Fuzzy C -Means Clustering (TSK)	0.247	245.31	0.120	0.347	0.725	0.474

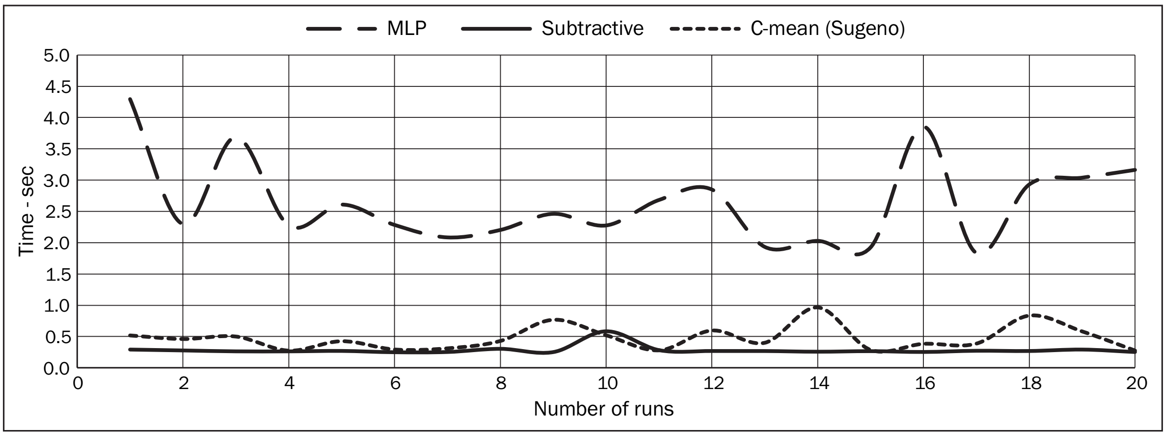

Figure 12 - Comparing the response time among the used prediction models

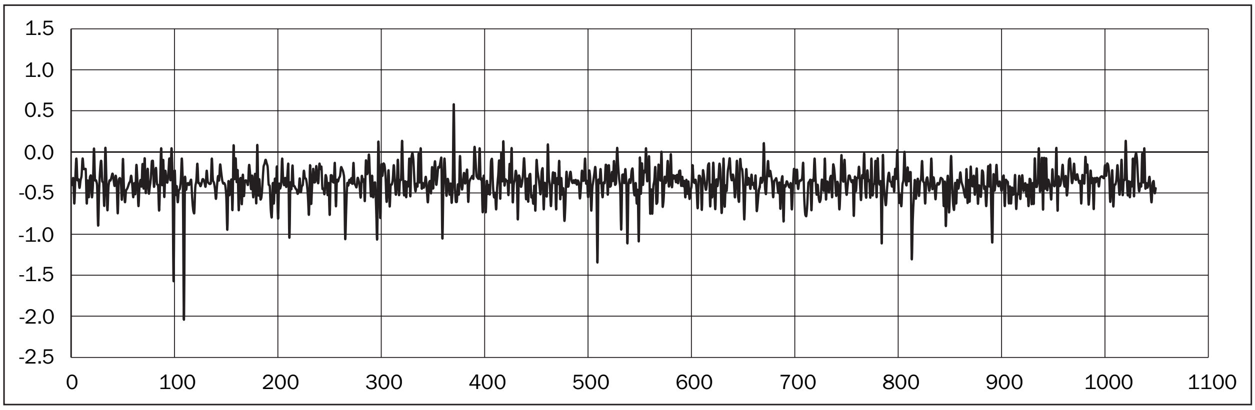

Figure 13 - The residuals for MLP model

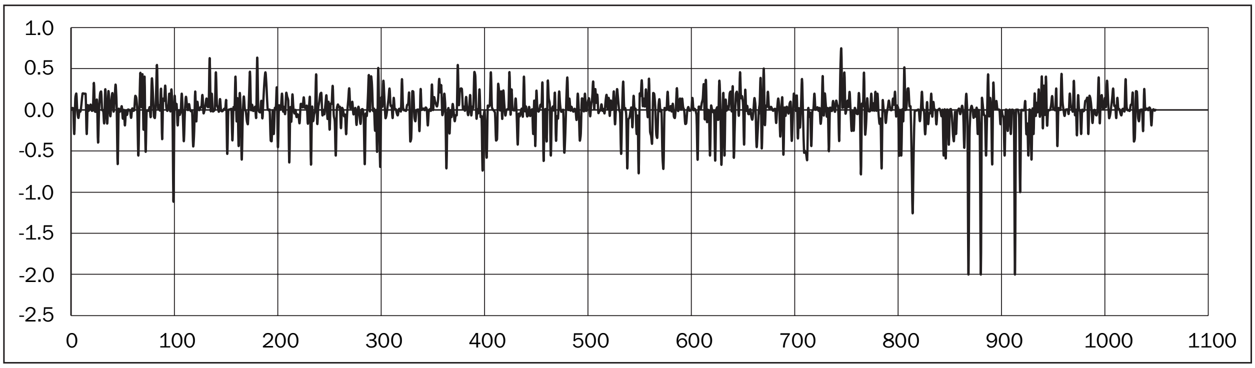

Figure 14 - The residuals for the FS clustering model

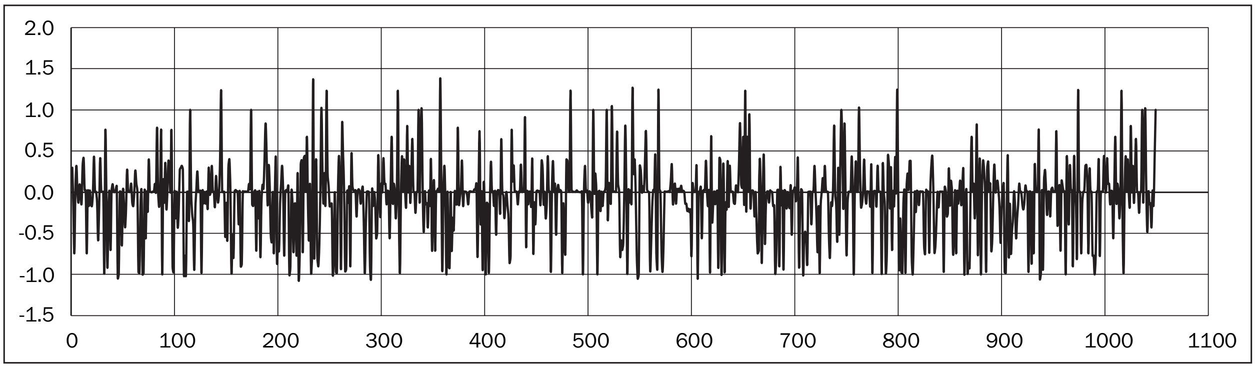

Figure 15 - The residuals for the FCM model

2. Twelve clusters were obtained from four clustering algorithms - hierarchical, K-means, FCM, and subtractive clustering - as the optimum number of clusters, as at this value, the mean silhouette coefficient and R-value converged in the clustering algorithms. Clustering should be applied to the input and output of the training records, which comprised approximately 800 records of the overall used data. The optimum number of clusters and the number of rules should be equal; therefore, 12 rules were created. In addition, each input and output was characterized by 12 membership functions. This number of clusters was applied to the FS clustering and FCM clustering.
3. Our procedure was able to identify the best two models based on precision (R) and response time (t). MLP model via exhaustive search took the greatest amount of time (2.635 seconds) with the best precision (R-value of 0.89). However, FS clustering took the least amount of time (0.284 seconds) with a precision with an R-value of 0.85. Thus, if a faster modelling time is desired, then FS clustering can be the right choice, but if precision is the goal, then MLP can be selected. In certain circumstances, MLP model could give higher accuracy, but MLP model would take more than 2.635/0.284 = 9.28 times longer to yield an answer than FS clustering. The comparison of multiple models in this research

provided a complete understanding of the relationship between input and output variables and allowed for identification of models yielding the highest prediction accuracy (MLP) and lowest response time (FS). The findings showed that more than one method can be suitable, depending on the selected criterion (precision and response time). While high precision results in better prediction levels of the crash severity, low response time can allow the developed system to assist agencies in performing real-time prediction with data from detectors and/or real-time traffic data.

4. The model adjusts itself by incorporating additional data, which means that determined models based on each criterion were modified with added data through induction procedure.

ACKNOWLEDGEMENT

The data have been collected with the cooperation of the local authority for traffic control in the framework of TSIP (traffic safety improvement program) funded by the EU. The authors are indebted to the local authority.

REFERENCES

[1] **Donnell, E.T.**, **Mason Jr. J.M.** *Predicting the frequency of median barrier crashes on Pennsylvania Interstate Highways*. Accident Analysis & Prevention 2006;38 (3):590–599.

[2] **Lord, D.** *Methodology for estimating the variance and confidence intervals of the estimate of the product of baseline models and AMFs*. Accident Analysis & Prevention 2008;40 (3):1013–1017.

[3] **Mussone, L.**, **Ferrari, A.** and **Oneta, M.** *An analysis of urban collisions using an artificial intelligence model*. Accident Analysis and Prevention, 1999;31 (6):705–718.

[4] **Zhang, J.**, **Lindsay, J.**, **Clarke, K.**, **Robbins, G.** and **Mao, Y.** *Factors affecting the severity of motor vehicle traffic crashes involving elderly drivers in Ontario*. Accident Anal. Prev. 2000;32 (1):117–125.

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

[6] **Yin, H.**, **Wong, S.C.**, **Xu, J.** and **Wong, C.K.** *Urban traffic flow prediction using fuzzy-neural network*, Transportation Research Part C 2002;10 (2):85–98.

[7] **Zhong, M.**, **Lingras, P.**, **Sharma, S.** *Estimation of missing traffic counts using factor, genetic, neural and regression techniques*. Transportation Research Part C 2004;12 (2):139–166.

[8] **Tong, H.Y.** and **Hung, W.T.** *Neural network modelling of vehicle discharge headway at signalized intersection: model descriptions and the results*, Transportation Research Part A 2002;36 (1):17–40.

[9] **Jin, X.**, **Cheu, R.L.** and **Dipti, S.** *Development and adaptation of constructive probabilistic neural network in freeway incident detection*. Transportation Research Part C 2002;10 (2):121–147.

[10] **Yuan, F.** and **Cheu, R.L.** *Incident detection using support vector machines*, Transportation Research Part C 2003;11 (3-4):309–328.

[11] **Subba Rao, P.V.**, **Sikdar, P.K.** and **Krishna Rao, K.V.** *Another insight into artificial neural networks through behavioral analysis of access mode choice*, Computers, Environment and Urban Systems 1998;22 (5):485–496.

[12] **Hensher, D.A.** and **Ton, T.T.** *A comparison of the predictive potential of artificial neural networks and nested logit models for commuter mode choice*. Transportation Research Part E 2000;36 (3):155–172.

[13] **Vythoulkas, P.C.** and **Koutsopoulos, H.N.** *Modelling discrete choice behavior using concepts from fuzzy set theory, approximate reasoning and neural networks*, Transportation Research Part C 2003;11 (1):51–73.

[14] **Mussone, L.**, **Rinelli, S.**, **Reitani, G.** *Estimating the accident probability of a vehicular flow by means of an artificial neural network*. Environment and Planning B: Planning and Design, 1996;23(6):667–675.

[15] **Sohn, S.** and **Lee, S.** *Data fusion, ensemble and clustering to improve the classification accuracy for the severity of road traffic accident in Korea*, Safety Science 2003;41(1):1–14.

[16] **Abdel-Aty, M.**, **Pande, A.**. *Identifying crash propensity using traffic speed conditions*. Journal of Safety Research 2005;36 (1): 97–108.

[17] **Abdelwahab, H.T.**, **Abdel-Aty, M.A.**. *Development of artificial neural network models to predict driver injury severity in traffic accidents at signalizes intersection*. Transportation Research Record 1746, Paper No. 01-2234, 2001;6–13.

[18] **Kunt, M.M.**, **Aghayan, I.** and **Noii, N.** *Prediction the traffic accident severity: comparing the artificial neural network, genetic algorithm, combined genetic algorithm and pattern search methods*. Transport 2011; 26(4): 353-366.

[19] **Zadeh, L.** *Fuzzy sets*. Information and Control 1965;8:338-353.

[20] **Akiyama, T.**, **Sho, C. F.** *Fuzzy mathematical programming for traffic safety planning on an urban expressway*. Transportation Planning and Technology 1993;17:179-190.

[21] **Hadji Hosseinlou, M.** and **Aghayan, I.** *Prediction of traffic accident severity based on fuzzy logic*, 8th International Congress on Civil Engineering, Shiraz, Iran. 2009;65p.

[22] **Kamijo, S.**, **Matsushita, Y.**, **Ikeuchi, K.**, **Sakauchi, M.** *Traffic monitoring and accident detection at intersections*. IEEE Transactions on Intelligent Transportation Systems 2000;1 (2):108–118.

[23] **Mussa, R.N.** and **Upchurch, J.E.** *Simulator evaluation of incident detection by transponder-equipped vehicles*. Transportation, 2002;29 (3):287–305.

[24] **Lanser, S.H.** and **Hoogendoorn, S.** *A fuzzy genetic approach to travel choice behavior in public transport networks*. CD-ROM Proc., Transportation Research Board 79th Annual Meeting, TRB, Washington, DC. 2000.

[25] **Niitymaki, J.** *General fuzzy rule base for isolated traffic signal control: rule formulation*. Transporation Planning and Technology, 2001;24 (3):237–247.

[26] **Ishak, S.S.** and **Al-Deek, H.M.** *Fuzzy art neural network model for automated detection of freeway incidents*. 1998;Transportation Research Record 1634:56–63.

[27] **Ruspini, E. H.** *A new approach to clustering*, Information and Control, 1969;15(1):22-32.

[28] **Dunn, J.C.** *A fuzzy relative of the ISODATA process and its use in detecting compact, well-separated clusters*. J. Cybernet, 1974;3:32–57.

[29] **Bezdek, J.** *Pattern Recognition with Fuzzy Objective Function Algorithms*, New York: Plenum Press;1981.

[30] **Sugeno M.** and **Yasukawa T.** *A fuzzy-logic-based approach to qualitative modelling*. IEEE Trans Fuzzy Syst 1993;1:7–31.

[31] **Chen JQ**, **Xi YG**, **Zhang Z.J.**, *A clustering algorithm for fuzzy model identification*. Fuzzy Sets Syst 1998;98:319–29.

[32] **Wang, X.**, **Wang, Y.** and **Wang, L.** *Improving fuzzy C-means clustering based on feature-weight learning*. Pattern Recognition Letters 2004;25:1123–1132.

[33] **Frigui, H.** and **Nasraoui, O.** *Unsupervised learning of prototypes and attribute weights*. Pattern Recognition, 2004;37:567–581.

[34] **Chiu, S.L.** *Fuzzy model identification based on cluster estimation*. Journal of Intelligent & Fuzzy System, 1994;2 (3).

[35] **Yager, R.R.** and **Filev, D.P.** *Approximate clustering via the mountain method*. IEEE Trans. Systems, Man, Cybernet. 1994;24 (8):1279–1284.

[36] **Hayajneh M.** and **Hassan A.** *Modelling the machinability of self-lubricated aluminium/alumina/graphite hybrid composites synthesized by the powder metallurgy method using a fuzzysubtractive clustering-based system identification method*. International Journal of Machining and Machinability of Materials 2008;3(3-4): 252–271.

[37] **MacQueen, J.B.** *Some methods for classification and analysis of multivariate observations*. Proceedings of Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1967;(1):281–297, Berkeley: University of California Press.

[38] **Fukunaga, K.** *Introduction to Statistical Pattern Recognition*. San Diego: Academic Press. 1990.

[39] **Pena, J.M.**, **Lozano, J.A.**, **Larranaga, P.** *An empirical comparison of four initialization methods for the K-means algorithm*. Pattern Recognition Letters, 1999;20:1027–1040.

[40] **Khan, S.** and **Ahmad, A.** *Cluster centre initialization algorithm for K-means clustering*. Pattern Recognition Letters 2004;25 (11):1293–1302.

[41] **Redmond, S.J.** and **Heneghan, C.** *A method for initializing the K-means clustering algorithm using kd-trees*. Pattern Recognition Letters 2007;28:965–973.

[42] **Sneath, P.H.** and **Sokal, R.R.** *Numerical Taxonomy*. London, UK:Freeman;1973.

[43] **King, B.** *Step-wise clustering procedures*. Journal of the American Statistical Association, 1967;69:86–101.

[44] **Guha, S.**, **Rastogi, R.** and **Shim, K.** *CURE: An efficient clustering algorithm for large databases*. In Proc. of 1998 ACM-SIGMOD Int. Conf. on Management of Data, 1998;73–84.

[45] **Guha, S.**, **Rastogi, R.** and **Shim, K.** *ROCK: A robust clustering algorithm for categorical attributes*. In Proc. of the 15th Int'l Conf. on Data Eng., 1999;512–521.

[46] **Karypis, G.**, **Han, E.H.**, and **Kumar, V.** *Chameleon: A hierarchical clustering algorithm using dynamic modelling*. IEEE Computer, 1999;32(8):68–75.

[47] **Wang, X.-Y.**, **Garibaldi, J. M.**, **Bird, B.** and **George, M. W.** *Novel Developments in Fuzzy Clustering for the Classification of Cancerous Cells Using FTIR Spectroscopy*. In: Valente de Oliveira, J. and Pedrycz, W., editors. Advances in Fuzzy Clustering and its Applications, John Wiley & Sons, Ltd, Chichester, UK. 2007

[48] **Hammouda, K.** and **Fakhreddine, K.** *A Comparative study of data clustering techniques*, SYDE 625: Tools of Intelligent Systems Design. Course Project. University of Waterloo, Ontario, Canada. 2002.