A geospatial neuro-fuzzy approach for identification of hazardous zones in regional transportation corridors

M. Effati[1], M. A. Rajabi[2,]*, F. Samadzadegan[3], Sh. Shabani[4]

Abstract

Today, one of the most detrimental consequences of developing road transportation systems in a country is traffic accident that places a huge financial burden on the society. The large number of passenger carriage together with high rate of crashes on highway corridors increase the necessity of a comprehensive study of regional transportation corridors hazardous zones. Current methods for identification of road segments with high potential of accident are mainly based on statistical approaches. Since there are not enough crash records for recently built roads, statistical methods are useless for this type of road networks. This paper presents a geospatial neuro-fuzzy approach for identification of hazardous zones on regional transportation corridors. It is a new research method for road hazardous zones modeling with abilities to infer meaning from complicated and ambiguous data. To demonstrate the framework, a prototype is developed and tested on Qazvin-Rasht (Iran) regional corridor. The results are compared against the existing black spots in the study corridor which Highway Police has determined based on statistical methods. Results show a correlation between the output of the proposed method and existing black spots. Moreover, the proposed approach has identified a few more zones in the corridor that were not determined by traditional statistical methods. The results confirms that this method is not only a cheaper one but also a robust means of analyzing the level of hazard associated with each road segment under consideration, specially when data are uncertain and incomplete.

Keywords: Road safety, Road hazardous zones (RHZ), Geospataial information system (GIS), Adaptive neuro fuzzy inference system (ANFIS).

1. Introduction

Transport systems have a major role in transporting goods and human, and it is obvious that existence of any problem in a transport system would endanger both the vehicles and travelers. One of these problems is traffic accident that places a huge financial burden on the society. For many classes of age groups, road traffic accident is among the leading causes of death. In particular, it is the first leading cause of death for people with the age of 15 to 29 years old [1]. The World Health Organization predicts that road collisions will jump from the ninth leading cause of death in 2004 to the fifth in 2030.

More than 80% of transportation in Iran is done through

road network [2], while the casualty statistics resulting from road traffic accidents is very high. To be more precise, recent investigations in Iran indicate that on average three persons per hour die in traffic accidents [3]. Statistics show that the number of road traffic accidents in Iran is still unacceptable specially when it is compared with the corresponding statistics for developed countries. In 2011 alone, more than 20,000 fatality accidents were recorded on Iran roads that caused considerable social and economical costs. According to a report by Iran Highway Police, most of the traffic accidents in 2011 (12232 fatalities) were occurred on regional transportation corridors.

One efficient approach to reduce crashes on regional roads is the identification of locations where traffic accidents tend to aggregate (the road hazardous zones or road black spots) through using the state of the art technologies and implementing corrective road safety measures. At present in Iran, accident data obtained from the "Analysis Form for Traffic Accidents" are used to identify the road segments with high potential for accident. This form is filled out by police officers for each traffic accident with injuries or deadly wounded casualties on a public road. Based on the information in these forms, hazardous road segments are identified using statistical approaches. Montella [4] has compared the performance of

* Corresponding author: Marajabi@ut.ac.ir

1 Ph.D. Candidate, Department of Geomatics Engineering, University of Tehran, Tehran, 14665-331, Iran
2 Assistant Professor, Departmet of Geomatics Engineering, University of Tehran, Tehran, 14665-331, Iran
3 Associate Professor, Department of Geomatics Engineering, University of Tehran, Tehran, 14665-331, Iran
4 Assistant Professor, Road Safety Department, Transportation Research Institute, Iran University of Science and Technology (IUST), Tehran, Iran

various statistical methods for hotspot identification.

Since there is no statistical information available for newly built corridors, therefore, statistical methods can hardly be used for the regional roads which are recently built. Consequently, there is a need for a general method by which road segments with high potential for accident on regional transportation corridors can be identified. Other limitation of the current methods is neglecting the spatial factor [5]. Geospatial Information Systems (GIS) can help to analyze and manage the effect of different environmental and roadway factors on concentration of traffic accidents. The application of GIS to transportation problems relies on the conventional functionality of GIS in terms of data management, visualization, query, and spatial analysis [6]. Moreover, a good understanding of the spatial and temporal distribution of accidents makes a considerable contribution to developing appropriate accident reduction programs and to evaluating their effectiveness [7].

Spatial analysis is used in this study for manipulating spatial data and visualizing relationships between spatial and non-spatial variables that are used for road hazardous zones identification. A major challenge in dealing with road hazardous zones using a geospatial method is considering experts linguistic judgments using an inference process. Therefore, the main objective of this study is to show the flexibility and advantages of using the neuro-fuzzy (ANFIS) and spatial analysis to regional corridors hazardous zones identification.

There are four main characteristics that this study takes into consideration: (a) presenting a general approach for identification of road hazardous zones in regional corridors through spatial analysis and neuro-fuzzy modeling (b) a fuzzy description of uncertain variables in road hazardous zones identification, (c) modifying imprecise model descriptions and improvement of initial fuzzy reasoning using the ANFIS learning capabilities, and (d) exploring environmental and roadway factors contributing to traffic accidents which leads to the identification of locations which are truly hazardous from a road safety authority perspective.

This paper is organized in six sections. After this introduction, a literature review is provided in section 2. The proposed methodology which consists of fuzzy reasoning and neuro-fuzzy concepts and its applicability to road hazardous zones identification is described in section 3. Section 4 presents the implementation process and an application of the proposed approach. Evaluating the results and discussion are discussed in section 5. The last section concludes with a discussion on the advantages of the proposed approach.

2. Literature Review

Various methods have been used in identifying road hazardous locations. In the simplest way, road hazardous zones are identified based on the total number of accidents. This is done by sorting locations in descending order based on traffic accident frequencies. Some studies [8-13] have used regression and Bayesian empirical for exploring the effect of roadway geometric and environment factors on highway corridors accident frequencies. Moreover, a variety of methods for point pattern analysis of traffic accidents have been proposed. These methods include quadrat analysis, nearest-neighbor distances, kernel density estimation, and K-function [7, 14-16]. Anderson [17] used a surface-based modeling approach through an interpolation function for road black spots identification. Yamada and Thill [7] demonstrated how the network version of a K-function method can be applied to spatial pattern analysis of traffic accidents. Yamada and Thill [18] studied the application of two types of local indicator of network-constrained clusters (the local Moran I statistic and the local Getis and Ord G statistic) on highway vehicle crashes. Steenberghen et al. [19] presented a method for identifying road hazardous locations in a network space. In their study, a road hazardous location is considered as a place where unexpected number of road accidents occurs under the assumption of total randomness of events.

The studies discussed above fail to consider the factors contributing to accidents that bias the process of road hazardous locations identification. Moreover, most of the current methods could not handle uncertain data in modeling road hazardous zones. Studying artificial intelligence (AI) based algorithms such as neuro-fuzzy system indicates that these algorithms with their unique abilities to infer meanings from ambiguous and deficient data are useful in many road safety applications [20-23]. Hadji Hosseinlou and Sohrabi [24] used a neuro-fuzzy inference system for predicting accident frequency on intercity roads. This study showed that the neuro-fuzzy modeling can estimate accident frequency in more than 96% of cases with a good quality.

On the other hand, due to the spatial nature of accidents, using geospatial based analytical methods is suitable for road hazardous locations analysis. Use of GIS for transportation purposes (GIS-T) is quickly becoming a mature domain of GIS technology application and has gained full recognition among transportation practitioners and academics [25]. Many studies have used GIS capabilities for analyzing safety and risk along roads, displaying crash locations and performing various crash analyses including analysis of traffic hazard intensity, preventing traffic accidents, and identifying road hazardous zones [26-30]. Carreker and Bachman [31] demonstrated that spatial analysis improves both the accuracy and efficiency of locating crashes along highway corridors. Steenberghen et al. [32] discussed the usefulness of spatial analysis and point pattern techniques for defining road accident black zones within urban agglomerations. Erdogan et al. [33] used GIS as a management system for accident analysis and determined hotspots on a highway corridor. This was done by converting accident records to a tabular form where they could then be matched to the road network.

In most of the studies discussed above, identification of road hazardous zone is not based on factors which contribute to accident occurrence and zones are identified if their accident occurrence is greater than a threshold. Also, a review on the scope of these studies indicates that

most researchers have chosen to analyze specific road geometry (such as curves or intersections) or specific vehicle types (such as large trucks). This study tries to develop a general approach for identifying road hazardous zones using spatial analysis and an inference process which considers experts linguistic judgments. It attempts to improve previous researches by integration of fuzzy reasoning and Artificial Neural Networks (ANN) learning capabilities in order to spatially identify road hazardous zones on regional highway corridors.

3. Methodology

3.1. Research design

This study presents a geospatial neuro-fuzzy approach for modeling hazardous zones and crash propensity characterization on regional highway corridors. It develops a methodology where roadway geometry and environmental factors are processed through an ANFIS to categorize the level of hazard associated with each road location under consideration. Proposed approach uses existing crash records and roadway information to calibrate and validate the model. Research methodology helps regional roads decision makers to determine which hazard factors are the most important ones, and ultimately to decide where hazard mitigation strategies should be employed. Fig. 1 shows the overall research design.
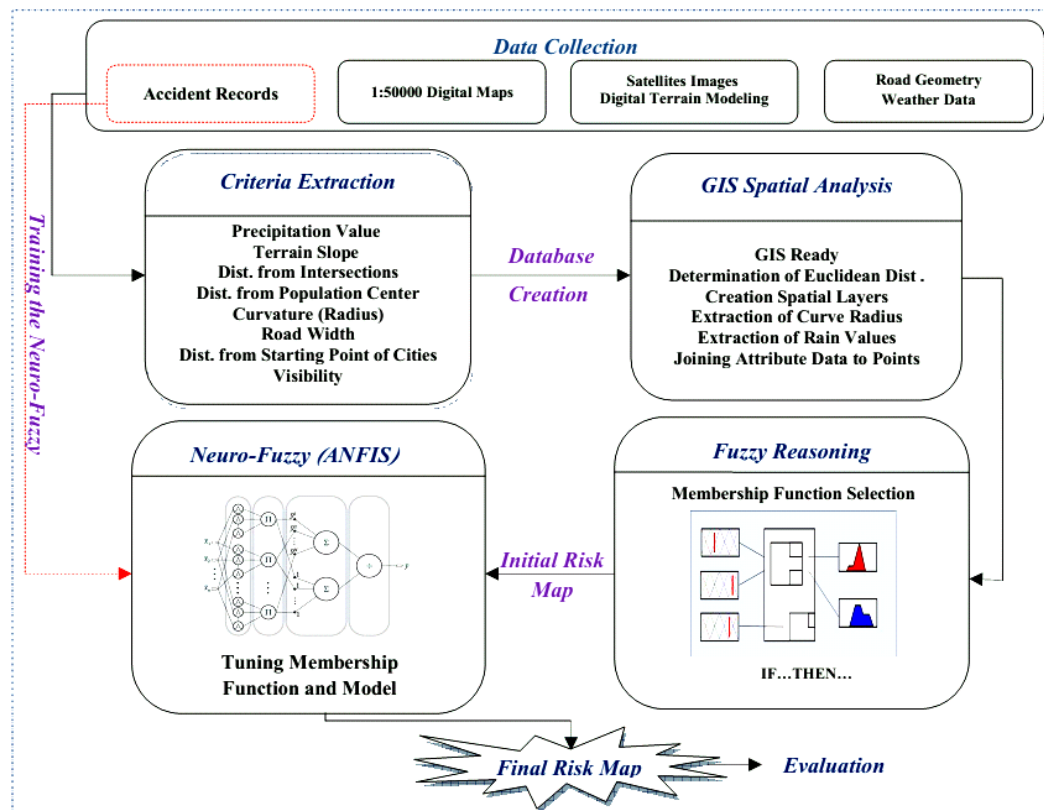

Fig. 1 Research methodology for identification of road hazardous zones

3.2. Fuzzy reasoning

Fuzzy logic is essentially a system for dealing with uncertainty and vagueness of concept [34]. Fuzzy set theory, introduced by Zadeh in the 1960s, resembles human reasoning in its use of approximate information and uncertainty to generate decisions [35]. Fuzzy logic allows objects to take partial membership in vague concepts. The main idea of fuzzy logic is that items in the real world are better described by having partial membership in complementary sets than by having complete membership in exclusive sets [35].

3.3. Hazardous zones identification based on fuzzy inference system

This research suggests the use of fuzzy inference system within the neural network for road hazardous zones identification because of the uncertainty associated with characterizing locations. Other reasons for using fuzzy logic in this research are:

– Research variables are continuous, imprecise, or ambiguous,

– Fuzzy logic is well suited for modeling continuous, real world systems,

- Fuzzy sets are an extension of crisp (two valued) sets to handle the concept of partial truth, which enables the modeling of uncertainties of natural language [36],
- Fuzzy logic has a tolerance for imprecision which can be exploited to achieve tractability, robustness, low solution cost, and better rapport with reality [37],

Traffic accidents are caused due to interaction of vehicle, driver, roadway and environmental factors. All these factors interact with each other and influence the occurrences and severity of accidents. In this study the criteria are selected based on possible causes identified in the accident reports, a review of the literature on road safety research, and the availability of data. These criteria include curvature, slope, visibility, distance from intersections, road width, distance from the starting point of roads, distance from population centers and weather condition (rain values). Table 1 illustrates these criteria and their descriptions.

Table 1 Description of criteria for road hazardous zones identification

Criteria	Description
Curvature	The shorter the radius the higher the hazard potential
Slope	Sections with higher slope have higher potential for hazard
Visibility	Sections with less visibility have higher potential for hazard
Distance from Intersection	Sections closer to intersections have higher hazard potential
Road Width	The narrower the road width is, the higher the hazard potential is
Distance from the Starting Point of roads (cities)	Sections closer to cities have higher hazard potential
Distance from Population Centers	Sections closer to the population centers have higher hazard potential
Rain Value	The higher the rain value is, the higher the hazard potential is

According to Table 2 these variables can be divided into two classes: road geometry design factors and environmental factors. Each of these variables is treated as a risk factor in road hazardous zones modeling. Fuzzy processing of hazard descriptors requires a specification of the linguistic labels which represent fuzzy sets. The linguistic variables and linguistic labels used for investigations of each geometry and environmental factors are listed in Table 2.

Table 2 Linguistic variables and labels for the fuzzy-based hazardous zones identification process

Type		Linguistic Variable	Linguistic Labels
Input	**Road Geometry Factors**	Curvature	Very Small, Small, Appropriate, High
		Slope	Low, Appropriate, High
		Visibility	Appropriate, Inappropriate
		Distance from Intersection	Very Near, Near, Far
		Road Width	Very Narrow, Narrow, Appropriate, Wide
		Distance From the Starting Point of Roads	Very Near, Near, Far
	Environmental Factors	Distance from Population Centers	Near, Moderate, Far
		Rain Value	Very Low, Low, High, Very High
Output	-	Danger	Absolutely Safe, Safe, Danger Prone , Danger, Very Danger

The type of fuzzy membership functions for each factor is very important, therefore, in this study various functions are tested and appropriate function for each variable is determined. Widely applied membership functions are bell-shaped and trapezoidal functions with maximum equal to 1 and minimum equal to 0. Trapezoidal functions are modeled with four parameters ($\alpha, \beta, \delta\alpha, \delta\beta$). This function is defined as (Fig. 2 and Eq. 1):
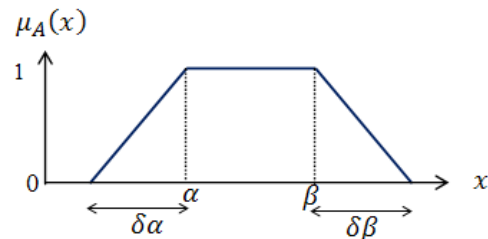

Fig. 2 Trapezoidal membership function

$$\mu_A\left(x\right)= \begin{cases} 1-\left(\alpha-x\right)/\delta\alpha & if\ \alpha-\delta\alpha \le x \le \alpha \\ 1 & if\ \alpha \le x \le \beta \\ 1-\left(\beta-x\right)/\delta\beta & if\ \beta \le x \le \beta+\delta\beta \\ 0 & otherwise \end{cases} \quad (1)$$

In special cases like symmetrical trapezoids and triangles the number of parameters is reduced to three. As the values of these parameters change, the membership functions vary accordingly, thus exhibiting various forms of membership functions [38]. An important aspect of these membership functions is that they are at least piecewise differentiable. This allows for tuning the membership function with a learning procedure that is discussed in the next section. This research uses the trapezoidal membership functions because of their simplicity, learning capability, and the short amount of time required for designing the system.

The main steps of proposed fuzzy inference system are: input, fuzzification, implication, aggregation and defuzzification. Fig. 3 shows the structure of fuzzy inference system for identification of road hazardous zones.
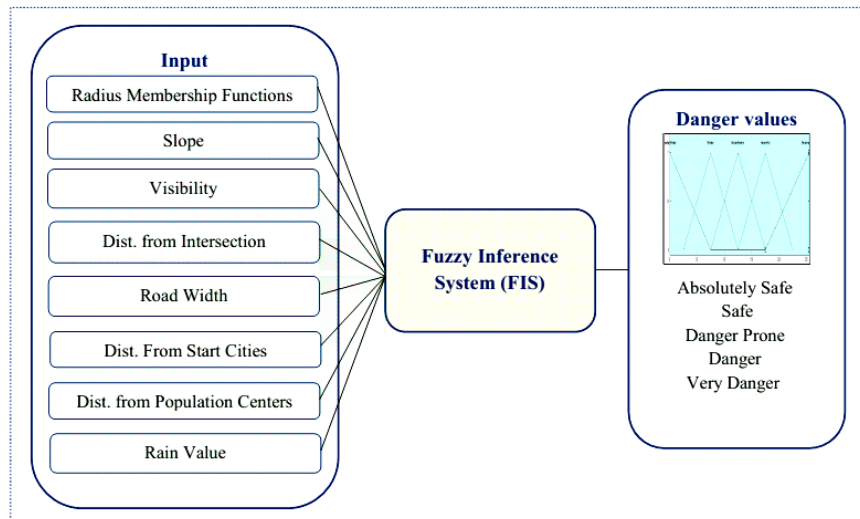

Fig. 3 Fuzzy inference system for identification of road hazardous zones

After implementing the criteria, in order to create a useful statement, complete sentences have to be formulated. Conditional statements, IF-THEN rules, are statements that make fuzzy logic useful. A single fuzzy IF-THEN rule can be formulated according to Eq. 2:

IF x is A; THEN y is B (2)

Where A and B are linguistic labels defined by fuzzy sets on the range of all possible values of x and y, respectively. The IF part of the rule "x is A" is called antecedent or premise, the THEN part of the rule "y is B" is called consequent. The antecedent is an interpretation that returns a single number between 0 and 1, whereas the consequent is an assignment that assigns the entire fuzzy set B to the output variable y. The antecedent may integrate several inputs using logical AND and OR. Fuzzy reasoning with fuzzy IF-THEN rules enables linguistic statements to be treated mathematically.

According to Table 2 this study restricts output of fuzzy process to five classes, namely absolutely safe, safe, danger prone, danger and very danger. Some samples of the IF-THEN fuzzy rules for determination of road hazardous zones have been given in Table 3.

Table 3 Some fuzzy rules

Sample Fuzzy Rules
1- IF *radius is* Very Small AND *slope* is High AND *visibility is* Inappropriate AND *distance from intersection is* Very Near AND *road width* is Very Narrow AND *rain value is* Very High AND *distance from cities* is Very Near THEN *point is* Very Dangerous.
3- IF *distance from population centers is* Near AND *radius is* Very Small AND *slope* is Appropriate AND *visibility is* Inappropriate AND *distance from intersection is* Very Near AND *road width* is Very Narrow AND *rain value is* Low AND *distance from cities* is Near THEN *point is* Very Dangerous.
4- IF *slope* is High AND *visibility is* Inappropriate AND *distance from intersection is* Far AND *road width* is Very Narrow AND *distance from cities* is Near THEN *point is* Dangerous.
5- IF *radius is* Very Small AND *visibility is* Inappropriate AND *distance from intersection is* Far AND *distance from cities* is Far THEN *point is* Dangerous.
6- IF *distance from population centers is* Near AND *radius is* Appropriate AND *slope* is Low AND

visibility is Appropriate AND *distance from intersection is* Far AND *road width* is Appropriate AND *rain value is* Low AND *distance from cities* is Far THEN *point is* Danger Prone.

7- IF *distance from population centers is* Moderate AND *radius is* Appropriate AND *slope* is Appropriate AND *visibility is* Appropriate AND *distance from intersection is* Near AND *road width* is Appropriate AND *rain value is* Low THEN *point is* Danger Prone.

8- IF *distance from population centers is* Far AND *radius is* High AND *slope* is Appropriate AND *visibility is* Appropriate AND *distance from intersection is* Far AND *road width* is Width AND *rain value is* High AND *distance from cities* is Far THEN *point is* Safe.

9- IF *distance from population centers is* Far AND *radius is* High AND *slope* is Appropriate AND *visibility is* Appropriate AND *distance from intersection is* Far AND *road width* is Width AND *rain value is* Very Low AND *distance from cities* is Far THEN *point is* Absolutely Safe.

This research uses the fuzzy Takagi and Sugeno (TSK) concept [39] for fuzzy based hazardous zones identification, because it offers some advantages with regard to computational efficiency and adaptive optimization. In TSK approach membership values in the premise part are combined with product inference to get the firing strength of each rule and the consequent part of each rule is modeled by a linear combination of the input variables plus a constant term (Eq. 3). The TSK rules can be expressed as following [40]:

$$R^j : \text{IF } x_1 \text{ is } A_1^j \text{ AND } x_2 \text{ is } A_2^j \text{ AND...AND}$$
$$x_n \text{ is } A_n^j , \text{THEN } f_j = a_0^j + a_1^j x_1 + K + a_n^j x_n \qquad (3)$$

Where R^j is the *jth* rule, $j = 1,2,. . .,m$, x_i is *ith* input variable, $i = 1,2,. . .,n$, A_i^j are linguistic terms of the premise part (e.g. Very Small, Small, Appropriate, High), f_j is the output variable (i.e. fuzzy indicator for the amount of hazard level), and a_i^j are coefficients of linear equations. The process of shaping the consequent (implication) is carried out and then aggregates the output fuzzy sets over all rules. The final output $\bar{y}$ (centroid defuzzifier) of hazardous zones identification fuzzy inference system is calculated by Eq. 4:

$$\bar{y} = \frac{\sum_{j=1}^{m} \mu_j f_j}{\sum_{j=1}^{m} \mu_j} \qquad (4)$$

where,

$$\mu_j = \mu_{A_1^j}(x_1).\mu_{A_2^j}(x_2)...\mu_{A_n^j}(x_n) \qquad (5)$$

The final output is the weighted average of the consequent equations rules. Fig. 4 shows the overall process of TSK fuzzy reasoning structure.
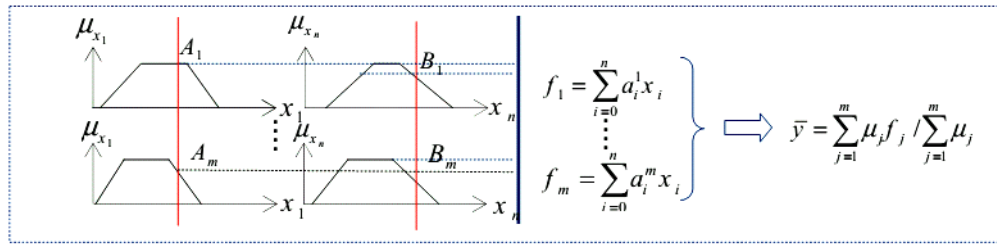

Fig. 4 Takagi and sugeno fuzzy reasoning structure

3.4. Improving model and tuning fuzzy membership functions

Fusion of artificial neural networks and fuzzy inference systems have attracted the growing interest of researchers in various scientific and engineering areas due to the growing need of adaptive intelligent systems to solve the real world problems [27]. A neuro-fuzzy system is a fuzzy system that is trained by a learning algorithm from neural network theory. This approach employs heuristic learning strategies derived from the domain of neural networks theory to support the development of a fuzzy system. The advantages of combination of neural networks and fuzzy inference systems are obvious. While the learning capability is an advantage from the viewpoint of fuzzy inference system, the automatic formation of linguistic rule base is another advantage from the viewpoint of neural network [41]. Although fuzzy logic can encode expert knowledge using linguistic labels, it usually takes a lot of time to tune the membership functions which quantitatively define these linguistic labels. Moreover, applications of fuzzy systems are restricted to the fields where expert knowledge is available and the number of input variables is small. Neural network learning techniques can automate this process and reduce development time and cost while improving performance and extracting fuzzy rules from numerical data

automatically (overcoming the fuzzy problem of knowledge acquisition). An integrated neuro-fuzzy reasoning system will possess the advantages of both neural networks (learning from examples, optimization of certain parameters) and fuzzy systems (meaningful representations, encoding knowledge, fuzzy rules and fuzzy reasoning).

The adaptive neuro-fuzzy inference system (ANFIS) is a specific type of neuro-fuzzy systems proposed by Jang (1993) [42]. ANFIS allows adaption of a model with hybrid learning rule and least squares error estimation. The architecture of ANFIS is based on a multilayer feed-forward network combined with a back-propagation gradient-descent-type learning algorithm that has a single output node [42]. ANFIS simulates Takagi–Sugeno–Kang fuzzy rule where the consequent part of the fuzzy rule is a linear combination of input variables and a constant. In the proposed fuzzy hazardous zones identification system each

Takagi and Sugeno rule comprises of two sets parameters: the premise parameters (α_i^j, β_i^j, $\delta\alpha_i^j$, $\delta\beta_i^j$), which are the parameters of the membership functions (Eq. 1), and the consequent parameters a_i^j, which are the coefficients of consequent part.

Fig. 5 shows the ANFIS structure which is proposed for identification of road hazardous zones. To reflect different adaptive capabilities, circle and square nodes are distinguished in this figure. A square indicates an adaptive node which has adjustable parameters whereas a circle represents a fixed node without parameters. The links between nodes in the network only indicate the flow direction without any weights. Input of the proposed ANFIS is divided into two groups (cf. Table 2), i.e., road geometry factors and environmental factors, and the output of the system is the identified hazard level.
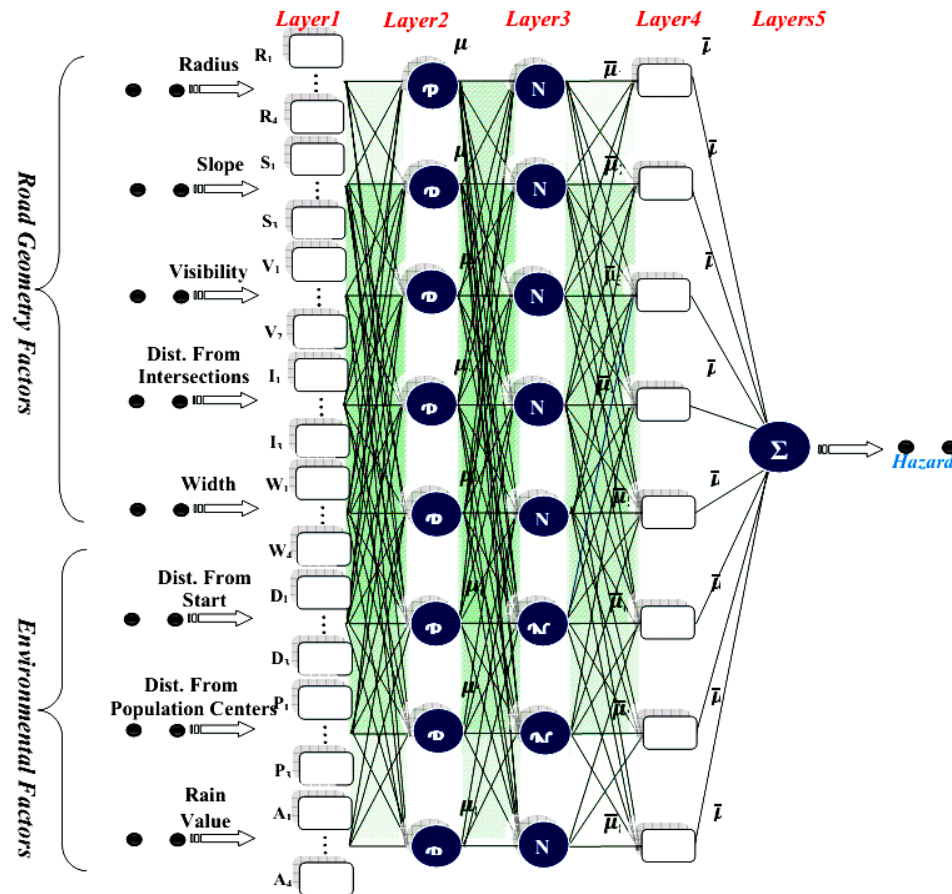

Fig. 5 Proposed ANFIS structure for road hazardous zones identification

Training process of this adaptive network is carried out in two steps, forward and backward. In the forward pass of the learning algorithm, processing proceeds up to Layer 4. In layer 4 the consequent parameters are adjusted and the network output in layer 5 indicates a certain hazard level. In the backward pass, the error rates propagate backward and the premise parameters in layer 1 are updated. In fact,

for the parameters in the layer 1, back-propagation algorithm is used. For training the parameters in the Layer 4, a variation of least-squares approximation or back-propagation algorithm is used; therefore, this system uses a hybrid learning algorithm in order to train the network. Characterizations of the node functions in each layer are explained below:

Layer 1: Is the fuzzification layer in which each node is as a membership function. Its output specifies the degree to which the given input x_i satisfies the linguistic terms A_i^j.

Layer 2: According to Eq. 5, the nodes in this layer multiply the incoming membership values and produce the firing strength of a rule (μ_j).

Layer 3: Is the normalization layer which normalizes the strength of all rules. In this layer the nodes labeled N determine the normalized firing strength of a rule ($\bar{\mu}_j = \mu_j \Big/ \sum_{j=1}^{m} \mu_j$).

Layer 4: Is the layer of adaptive nodes; in this layer each node j calculates the weighted consequent values ($\bar{\mu}_j f_i$).

Layer 5: The single node in this output layer receives the final result of ANFIS system (summation of the network outputs of the nodes in Layer 4).

4. Implementation and Results

This section briefly explains the implementation steps of the proposed method including data acquisition and preparation, modeling hazardous zones based on fuzzy inference system, optimization of the model and finally the evaluation of results.

4.1. Data and study area

To illustrate how the proposed methodology works, part of Qazvin-Rasht highway corridor which connects Tehran to the North of Iran (Gilan), is chosen as the study corridor. The study corridor is located in a mountainous region where elevation ranges from approximately 300 to 2394 m. It is a two-lane two-way road and according to previous accident records, some road segments have high potential for accidents. Fig. 6 shows the study area.

This study uses several sources of primary data in various formats. Table 4 describes the primary data of this research.
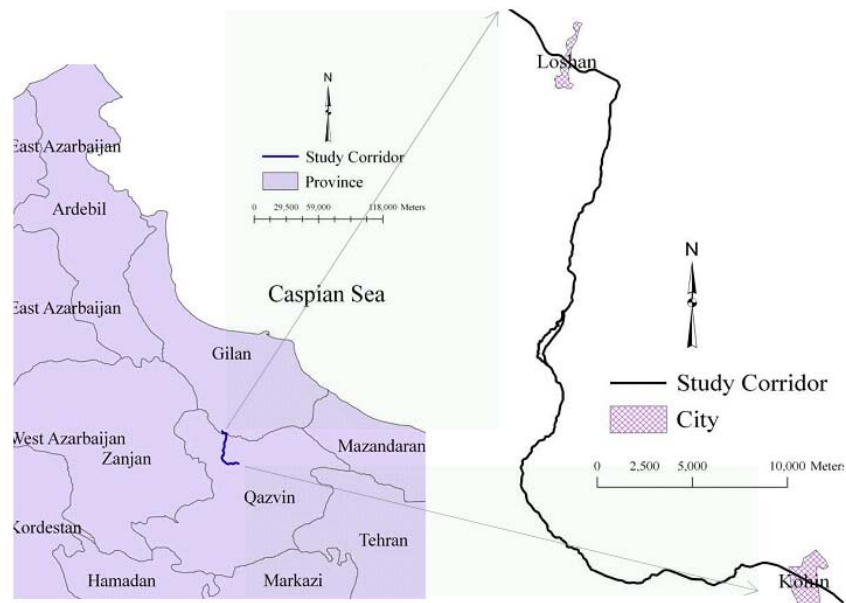

Fig. 6 The study corridor

Table 4 Description of primary data layers.

Data Layers	Source	Scale/Resolution	Description
Topographic Map	National Cartographic Centre (NCC)	1: 50000	**Digital**
Digital Elevation Model	National Cartographic Centre	10 meter	**Digital**
Geometric Specification	Road Ministry	-	**Attribute**
Weather Stations Information	Meteorological Organization	-	**To acquire rain values along the road**
Crash Data	Highway Police	-	**-**
Excising Hazardous Zones	**Ministry of Road and Urban Development**	-	**For evaluation and test**

4.2. Experimental investigations

For testing the proposed method on the study corridor, it is divided into smaller segments using dynamic segmentation. Next, a database of all data and layers is generated and road geometry and environmental attributes are assigned to each segment. Each variable in Table 1 is treated as a risk factor in analysis of hazardous zones identification where critical standard boundaries for each criterion (observed indicators) are determined. Since classes or groups of data with boundaries are not sharply defined, their indicators and relationships have uncertain definition. Therefore, some uncertainties are lying in this method. Fuzzy set theory and linguistic variables is a useful tool for solving uncertainty. It also facilitates subsequent integration of data layers in the generation of composite risk maps.

Prior to fuzzy process the membership functions of each factor have to be specified. The initial membership functions are depicted in Fig. 7.
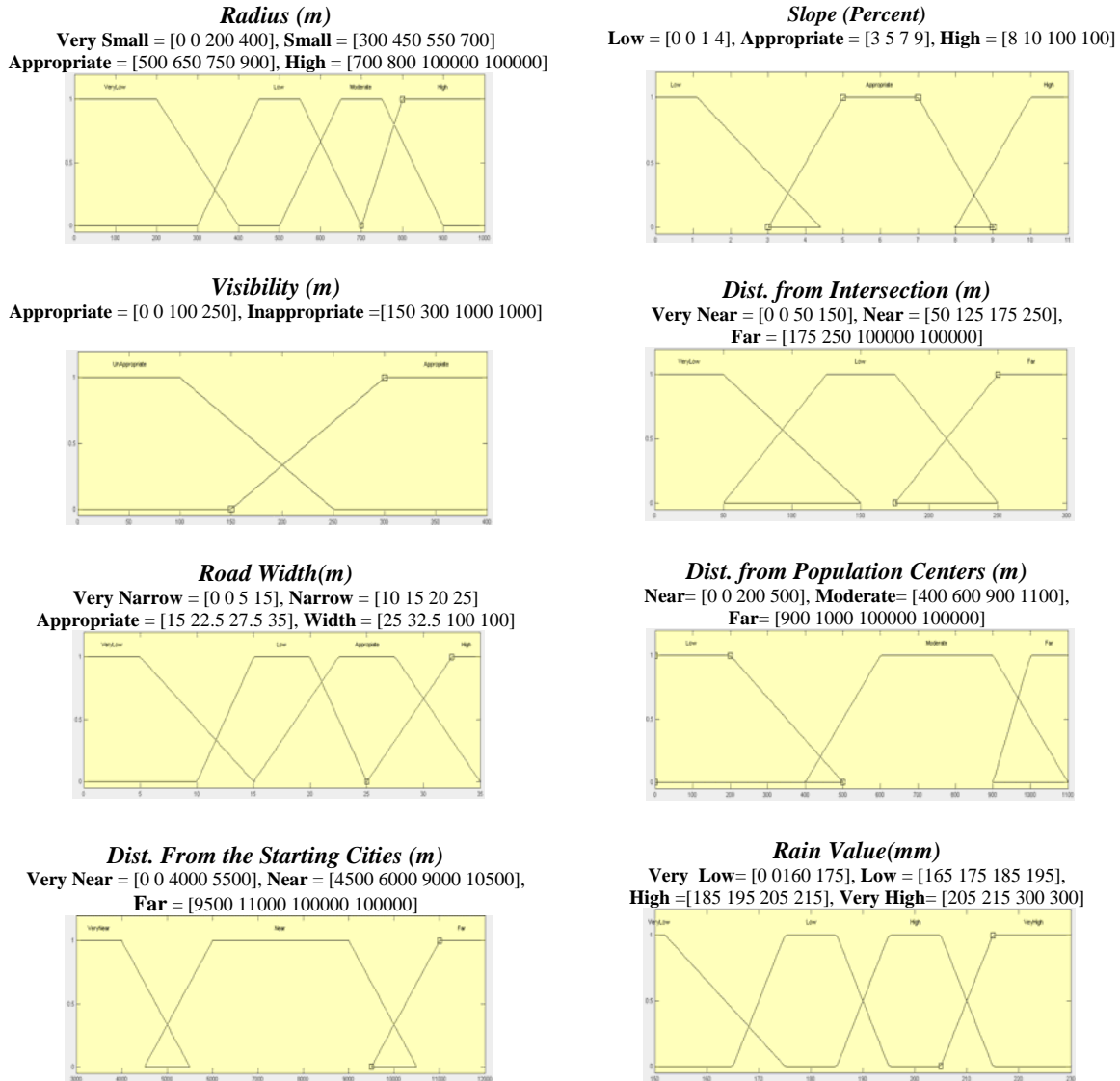
Radius (m)
Very Small = [0 0 200 400], **Small** = [300 450 550 700]
Appropriate = [500 650 750 900], **High** = [700 800 100000 100000]

Slope (Percent)
Low = [0 0 1 4], **Appropriate** = [3 5 7 9], **High** = [8 10 100 100]

Visibility (m)
Appropriate = [0 0 100 250], **Inappropriate** =[150 300 1000 1000]

Dist. from Intersection (m)
Very Near = [0 0 50 150], **Near** = [50 125 175 250],
Far = [175 250 100000 100000]

Road Width(m)
Very Narrow = [0 0 5 15], **Narrow** = [10 15 20 25]
Appropriate = [15 22.5 27.5 35], **Width** = [25 32.5 100 100]

Dist. from Population Centers (m)
Near= [0 0 200 500], **Moderate**= [400 600 900 1100],
Far= [900 1000 100000 100000]

Dist. From the Starting Cities (m)
Very Near = [0 0 4000 5500], **Near** = [4500 6000 9000 10500],
Far = [9500 11000 100000 100000]

Rain Value(mm)
Very Low= [0 0160 175], **Low** = [165 175 185 195],
High =[185 195 205 215], **Very High**= [205 215 300 300]

Fig. 7 Initial input membership functions

Formulation of the fuzzy rules requires exact consideration of impact of each descriptor on the accident occurrence and hazard values, complexity of each factor and the experience of experts. Therefore, selection of appropriate rules for road hazardous zones identification is a sensitive and important subject.

After defining the input and output of fuzzy inference system and its membership functions and rules, the value of danger for each segment is determined. Danger values in the proposed fuzzy inference system are classified in the range of 0 (absolutely safe) to 250 (very dangerous). Fig. 8 depicts the output of applying fuzzy reasoning on the study

corridor for identification of zones with high potential for accident. In this figure each point indicates a hazardous

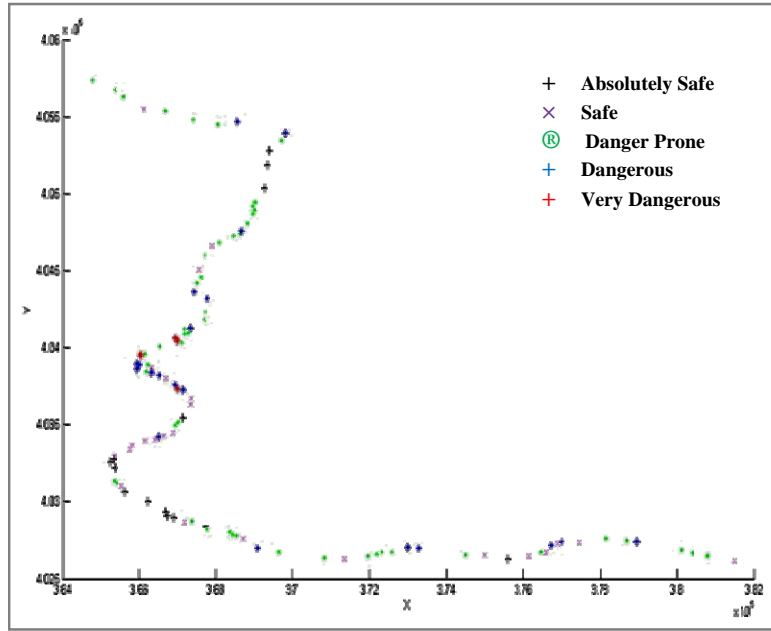zone and each danger class is shown with a special symbol and color.

Fig. 8 The study corridor hazardous zones (approach: fuzzy reasoning)

4.2.1. Hazardous zones identification based on neuro-fuzzy modeling

Definition of membership functions for the hazard factors is an important and complex problem. This section tries to use a neuro-fuzzy inference system to tune and improve the fuzzy membership functions and fuzzy model through a training process. For the training process of the proposed neuro-fuzzy inference system a very large number (100) of training datasets with appropriate distribution along the corridor is used. Training data are selected by considering the transportation experts knowledge and the corridor crash records in the past. For verification of learning process, thirty extra sample points which are not included in the objective function of neuro-fuzzy optimization process, are used.

The output of the proposed reasoning system (y) is calculated using the training dataset and employing Eqs. 4 and 5. Assuming that y^d is the desired output for input dataset, an error measure can be defined by the squared error between the actual output (y) and the desired output (Eq. 6):

$$E = (y - y^d) = [\left(\frac{\sum_{j=1}^{m} \mu_j f_j}{\sum_{j=1}^{m} \mu_j} \right) - y^d] \tag{6}$$

According to Eq. 3 and 6 the error measure E depends on the tuning parameters ($\alpha_i^j, \beta_i^j, \delta\alpha_i^j, \delta\beta_i^j$) and a_i^j. Parameter update starts with an initial setting of the values

of the premise parameters and consequent parameters. The fuzzy reasoning steps are repeated with the update of the consequent and premise parameters until the change in the inference error is less than a predefined threshold and this indicates convergence of the process. Fig. 9 shows the error measure (y axis) as a function of the number of iterations (x axis) for checking and training data.
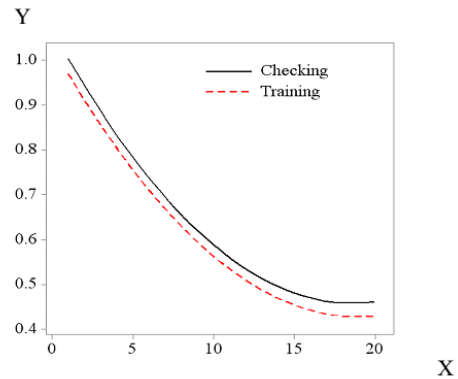

Fig. 9 Training process

Fig. 9 shows that the checking process curve is, as expected, slightly above the error curve obtained for the training data. The very small difference for the final iterations indicates that system is correctly trained. Fig. 10 shows the adapted membership functions after the training process.

Comparing the initial membership functions in Fig. 7 with the tuned ones in Fig. 10 indicates that after training most of the membership functions have changed significantly. This expresses that the training process of

M. Effati, M. A. Rajabi, F. Samadzadegan, Sh. Shabani

neuro-fuzzy system is quite effective.

Fig. 11 depicts the output of applying proposed ANFIS model on the study corridor for identification of zones with high potential for accident. Comparing Fig. 11 and Fig. 8 indicates that by changing membership functions in the training process the values of hazard in some points have changed.
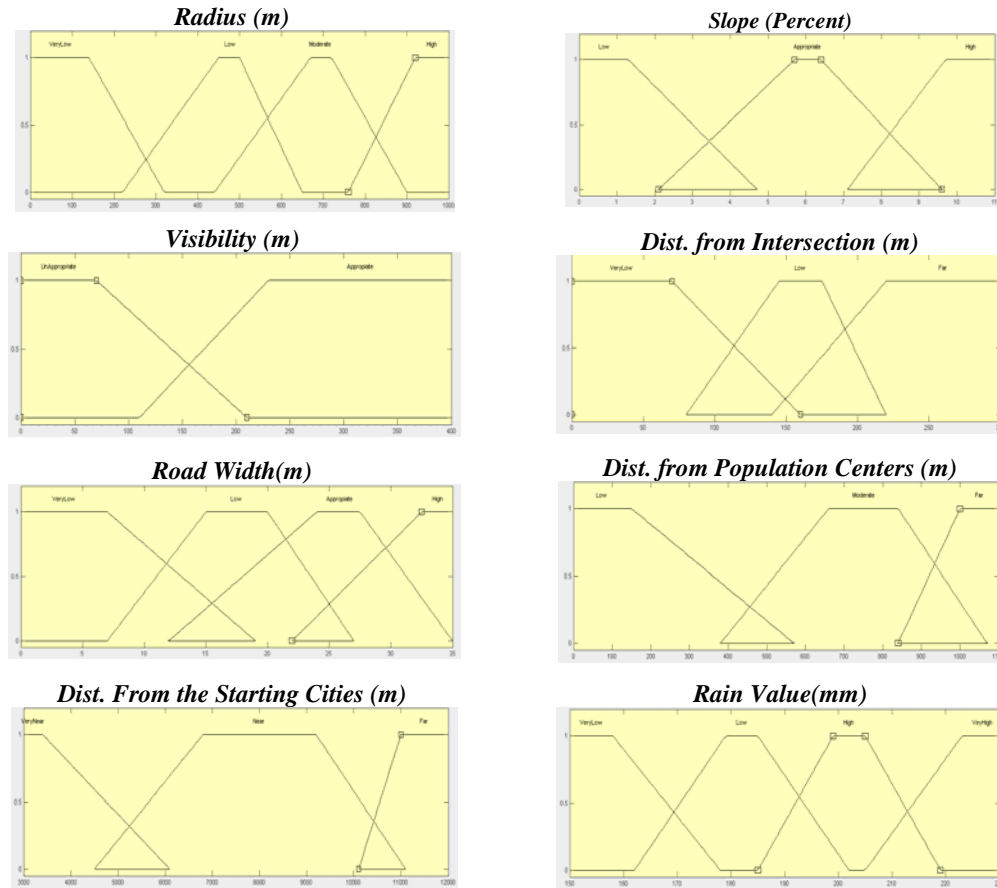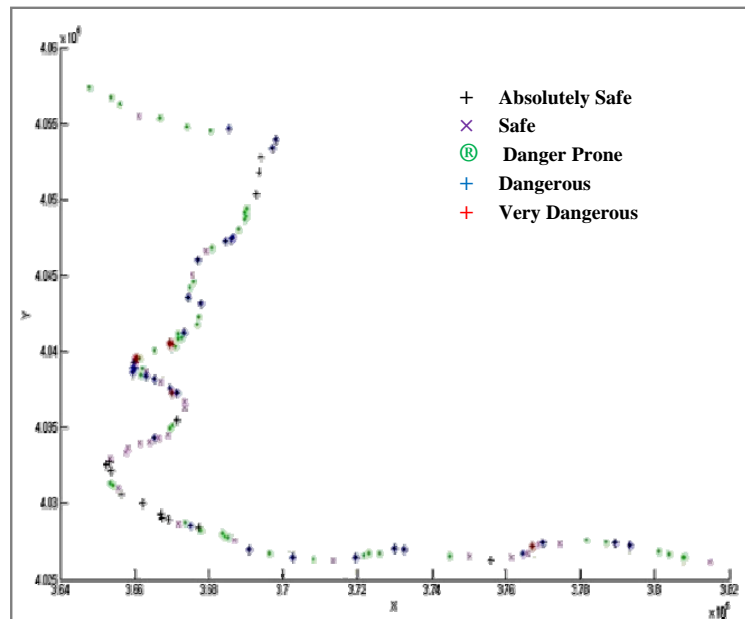

Fig. 10 Tuned membership functions

Fig. 11 The study corridor hazardous zones (approach: anfis model)

5. Discussion

This study proposed an ANFIS model for identification of regional transportation corridors hazardous zones with the ability to deal with imprecise, uncertain or ambiguous datasets and knowledge. It can consider the relationships among input data and hazard value which are difficult to handle using conventional regression and logic methods. In comparison with other studies that used statistical methods for road hazardous location identification [4, 7, 13], the proposed method is a non-parametric technique which does not require prior knowledge of the crash factors. Furthermore, according to Montella's study [43] such artificial intelligence based approaches can consider conditional interactions among input data.

Using the proposed ANFIS model, the study corridor hazardous zones are obtained and visualized in a map (Fig. 12). In this map red and blue points indicate locations of very dangerous and dangerous zones, respectively. Table 5 illustrates the specifications of hazardous zones.
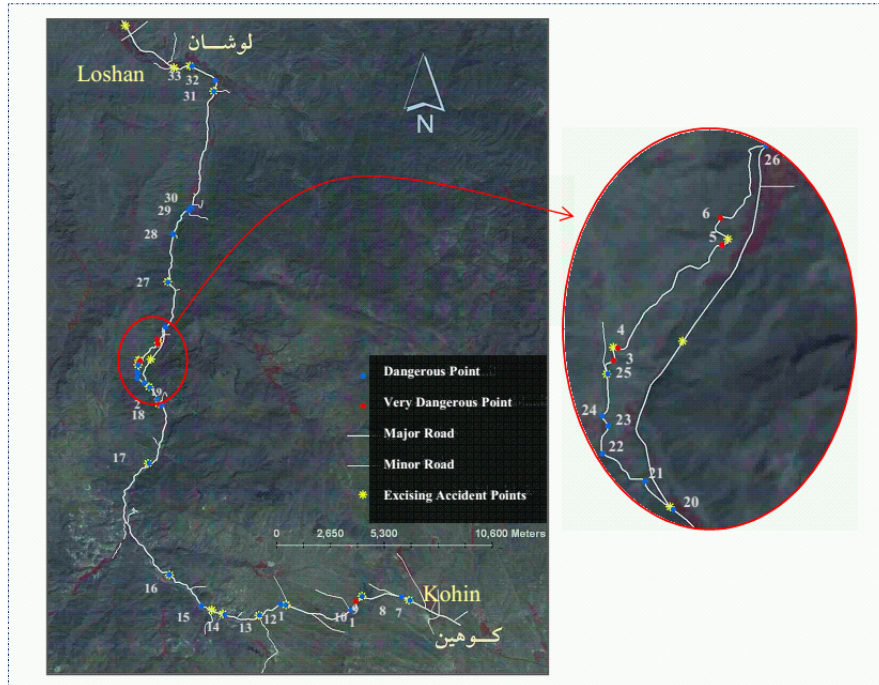

Fig. 12 Map of hazardous zones along the study corridor (red: very dangerous, blue: dangerous, yellow: excising accident zones)

Table 5 Kohin–Loshan hazardous zones specifications

Zone	Danger Value	Distance from the Start (km)	Start of Route	End of Route
1	Very Dangerous	4-5	Nasim-Shomal Intersection	Loshan
2	Very Dangerous	31-32	Nasim-Shomal Intersection	Loshan
3	Very Dangerous	34-35	Nasim-Shomal Intersection	Loshan
4	Very Dangerous	34-35	Nasim-Shomal Intersection	Loshan
5	Very Dangerous	35-36	Nasim-Shomal Intersection	Loshan
6	Very Dangerous	36-37	Nasim-Shomal Intersection	Loshan
7	Dangerous	1-2	Nasim-Shomal Intersection	Loshan
8	Dangerous	2-3	Nasim-Shomal Intersection	Loshan
9	Dangerous	3-4	Nasim-Shomal Intersection	Loshan
10	Dangerous	4-5	Nasim-Shomal	Loshan

			Intersection	
11	Dangerous	5-6	Nasim-Shomal Intersection	Loshan
12	Dangerous	5-6	Nasim-Shomal Intersection	Loshan
13	Dangerous	7-8	Nasim-Shomal Intersection	Loshan
14	Dangerous	9-10	Nasim-Shomal Intersection	Loshan
15	Dangerous	11-12	Nasim-Shomal Intersection	Loshan
16	Dangerous	14-15	Nasim-Shomal Intersection	Loshan
17	Dangerous	27-28	Nasim-Shomal Intersection	Loshan
18	Dangerous	31-32	Nasim-Shomal Intersection	Loshan
19	Dangerous	31-32	Nasim-Shomal Intersection	Loshan
20	Dangerous	33-34	Nasim-Shomal Intersection	Loshan
21	Dangerous	33-34	Nasim-Shomal Intersection	Loshan
22	Dangerous	33-34	Nasim-Shomal Intersection	Loshan
23	Dangerous	34-35	Nasim-Shomal Intersection	Loshan
24	Dangerous	34-35	Nasim-Shomal Intersection	Loshan
25	Dangerous	34-35	Nasim-Shomal Intersection	Loshan
26	Dangerous	38-39	Nasim-Shomal Intersection	Loshan
27	Dangerous	44-45	Nasim-Shomal Intersection	Loshan
28	Dangerous	57-58	Nasim-Shomal Intersection	Loshan
29	Dangerous	58-59	Nasim-Shomal Intersection	Loshan
30	Dangerous	45-46	Nasim-Shomal Intersection	Loshan
31	Dangerous	59-60	Nasim-Shomal Intersection	Loshan
32	Dangerous	59-60	Nasim-Shomal Intersection	Loshan
33	Dangerous	61-62	Nasim-Shomal Intersection	Loshan

In Fig. 12 yellow points show the existing black spots along the corridor which Highway Police has determined based on statistical methods. These spots are compared with the hazardous zones that are identified using the proposed model. The results show a good correlation between existing black spots (yellow dots) and the hazardous zones (red and blue points). This confirms Hadji Hosseinlou and Sohrabi's study [24] which showed applicability of neuro-fuzzy systems in predicting traffic hot spots in intercity roads.

In some instances the zones of highest risk do not exactly overlay with black spots. Moreover, there are some hazardous zones that were not determined using traditional statistical methods and are not in Highway Police database. The authors believe these zones could be prone to accident in the future and should be taken into consideration in the study corridor safety programs.

For quality control of the purposed approach as well as verification of the spatial analysis performance, the calculated hazardous zones (which their centers are within 20 m of the existing black spots) are compared with existing black spots. Table 6 lists the coordinate differences of existing and calculated ones along x and y direction. Moreover, using Eq. 7 values of RMSE are calculated.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_i - X'_i)^2}{n-1}} \qquad (7)$$

where X'_i is the locations of existing black spots, X_i is the centers of calculated zones, and n is number of correspond zones.

Table 6 Evaluation of results

Points	Difference	
	X(m)	Y(m)
2	1	8.3
4	9	11.1
7	1.2	5.4
9	3.4	8.9
11	6.5	6
13	7.4	3.9
14	15	8.9
16	5.8	9.1
17	8	2.4
20	7.8	8
25	2	1.7
27	6.8	4.9
31	4	3.9
33	12.2	5
RMSE	7.51	6.81
Total RMSE	10.13	

RMSE value shows that using a 10.13 m threshold the corresponding points lay in the same segment. Several factors may explain this small differences including error associated with the crash data, approximate determination of existing black spots, error in road geometry and environmental data, temporary obstructions in the corridor, and other parameters and factors that are unaccounted in this analysis.

Unlike statistical approaches that consider the frequency of accidents for identifying black spots, the proposed method can assess how a variation in one or more of input factors can affect the danger level on hazardous zones. Recognizing the most contributing factors and their effect on concentration of accidents helps not only to reduce the number of crashes on regional corridors but also to carry out precautionary safety operations on hazardous zones. Since, sensitivity analysis is done on proposed model by changing various input factors in ANFIS model, gradually adding and changing the ranges of the membership functions. Consequently, it is discovered that road curvature, slope and proximity to intersections tend to produce statistically significant differences in results. Therefore, these factors are the most important factors which affect the danger level on hazardous zones and need to be addressed in future hazard mitigation strategies in the study corridor.

6. Conclusions

This study developed a geospatial neuro-fuzzy approach for modeling hazardous zones in regional transportation corridors. It used fuzzy logic to model transportation experts' linguistic judgments in its imprecise and vague nature. Spatial analysis was used for manipulating spatial data and visualizing the relationships between spatial and non-spatial variables related to hazardous zones. An important aspect of the proposed approach was improving the fuzzy model using an optimization process through which membership functions of the linguistically formulated fuzzy sets are tuned. Moreover, through the proposed ANFIS model it is possible to show how a variation in one or more of input factors can affect the danger level on hazardous zones. The study used past crash records along with roadway information to calibrate and validate the model. It was tested on a regional highway corridor and the calculated hazardous zones were compared with the existing black spots which were obtained using statistical approaches. It was found a good correlation between existing black spots and the calculated hazardous zones. Furthermore, the approach explored some hazardous zones in the study corridor that were not determined using statistical methods.

As a future work the procedure used to identify hazardous zones on regional transportation corridors could be integrated with other road safety and crash analysis programs and/or applied to other types of infrastructure (rural roads, freeways, etc.).

References

[1] World Health Organization Global (WHO). Status report on road safety, 2009

[2] RMTO I.R. of Iran Road Maintenance and Transportation Organization; Annual Report, (in Farsi).<http://www.rmto.ir/NewTTO/DynaStat/DynaStatF/> (21.02.14). 2012.

[3] FMOI, Forensic Medicine Organization of Iran, Statistical Data, Accidents, (in Farsi), <http://www.lmo.ir/?siteid=1&pageid=1347> (19.02.13), 2009.

[4] Montella AA. Comparative analysis of hotspot identification methods, Accident Analysis and Prevention, 2009, No. 2, Vol. 42, pp. 571–581.

[5] Siddiqui C, Abdel-Aty M, Huang H. Aggregate nonparametric safety analysis of traffic zones, Accident Analysis and Prevention, 2012, Vol. 45, pp. 317-325.

[6] Thill JC. Transportation applications of geographic information systems, Invited contribution to the Manual of Geographic Information Systems, M. Madden, Editor, ASPRS, Washington, D.C, 2009, pp. 1035-1049.

[7] Yamada I, Thill JC. Comparison of planar and network K-functions in traffic accident analysis, Journal of Transport Geography, 2004, Vol. 12, pp. 149-158.

[8] Jovanis P, Delleur J. Exposure-based analysis of motor vehicle accidents, Transportation Research Record, 1983, Vol. 910, pp. 1-7.

[9] Miaou S, Lum H. Modeling vehicle, accidents and highway geometric design relationships, Accident Analysis and Prevention, 1993, No. 6, Vol. 25, pp. 689-709.

M. Effati, M. A. Rajabi, F. Samadzadegan, Sh. Shabani

[10] Knuiman M, Council F, Reinfurt D. The effect of median width on highway accident rates, Transportation Research Record, 1993, Vol. 1401, pp. 70-80.

[11] Miaou SP. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions, Accident Analysis and Prevention, 1994, No. 4, Vol. 26, 26, pp. 471-482.

[12] Shankar V, Mannering F, Barfield W. Effect of roadway geometric and environment factors on rural freeway accident frequencies, Accident Analysis and Prevention, 1995, No. 3, Vol. 27, pp. 37-89.

[13] Cheng W, Washington SP. Experimental evaluation of hotspot identification methods, Accident Analysis and Prevention, 2005, No. 5, Vol. 37, pp. 870–881.

[14] Cressie N. Statistics for Spatial Data, John Wiley & Sons, New York, 1993.

[15] Diggle PJ. Spatial Analysis of Spatial Point Patterns, Academic Press, New York, 1983.

[16] Fotheringham A, Brunsdon C, Charlton M. Quantitative geography: perspectives on spatial data analysis, London, Sage Publication, 2000.

[17] Anderson TK. Kernel density estimation and K-means clustering to profile road accident hotspots, Accident Analysis and Prevention, 2009, No. 3, Vol. 41, pp. 359-364.

[18] Yamada I, Thill JC. Local indicators of network-constrained clusters in spatial patterns represented by a link attribute, Annals of the Association of American Geographers, 2010, No. 2, Vol. 100, pp. 1-17.

[19] Streenberghen T, Aerts K, Thomas I. Spatial clustering of events on a network, Journal of Transport Geography, 2010, No. 3, Vol. 18, pp. 411-418.

[20] Busch F, Cremer M. Ghio A, Henninger T. A muti-model approach with fuzzy reasoning for traffic state estimation and incident detection on motorways, In: ERTICO (Ed), Proceedings of the 1st world congress on applications of Transport Telematics and Intelligent Vehicle Highway Systems, Boston and London, 1995, pp. 145-1252.

[21] Pappis C, Mamdani E. A fuzzy controller for a traffic junction, Institute of Electrical and Electronics Engineers (IEEE), Transactions on systems, man and cybernetics, 1977, No. 17, Vol. 7, pp. 707-717.

[22] Sayed T, Abdelwahab W, Navin F. Identifying the accident-prone locations using fuzzy pattern recognition, Journal of Transportation Engineering, 1995, No. 4, Vol. 121, pp. 352-358.

[23] Adeli H, Jiang X. Neuro-fuzzy logic model for freeway work zone capacity estimation, Journal of Transportation Engineering, 2003, No. 5, Vol. 129, pp. 484-493.

[24] Hadji Hosseinlou M, Sohrabi M. Predicting and identifying traffic hot spots applying neuro-fuzzy systems in intercity roads, International Journal of Environmental Science and Technology, 2009, No. 2, Vol. 6, pp. 09-314.

[25] Thill JC. Geographic Information Systems in Transportation Research, Pergamon, U.K, Oxford, 2001.

[26] Jha M, McCall C, Schonfeld P. Using GIS, genetic algorithms and visualization in highway development, Computer-Aided Civil and Infrastructure Engineering, 2001, No. 6, Vol. 16, pp. 399 - 414.

[27] Fuller D, Jeffe M, Williamson RA, James D. Satellite remote sensing and transportation lifelines: safety and risk analysis along rural southwest roads, ISPRS Commission I, Pecora, 2002, Vol. 15.

[28] Gundogdu IB. Applying linear analysis methods to GIS-supported procedures for preventing traffic accidents: Case study of Konya, Safety Science, 2010, No. 6, Vol. 48, pp. 763-769.

[29] Ha HH, Thill JC. Analysis of traffic hazard intensity: a spatial epidemiology case study of urban pedestrians, Computers. Environment and Urban Systems, 2011, No. 3, Vol. 35, pp. 230-240.

[30] Effati M, Rajabi MA, Samadzadegan F, Blais JAR. Developing a novel method for road hazardous segment identification based on fuzzy reasoning and GIS, Journal of Transportation Technologies, 2012, No. 1, Vol. 2, pp. 32-40.

[31] Carreker LE, Bachman W. Geographic information system procedures to improve speed and accuracy in locating crashes, Transportation Research Record, TRB, National Research Council, Washington D.C, 2000, Vol. 1719, pp. 215-218.

[32] Steenberghen T, Dufays T, Thomas I, Flahaut B. Intra-urban location and clustering of road accidents using GIS: a Belgian, International Journal of Geographical Information Science, 2004, No. 2, Vol. 18, pp. 169- 181.

[33] Erdogan S, Yilmaz I, Baybura T, Gullu M. Geographical information systems aided traffic accident analysis system case study: city of Afyonkarahisar, Accident Analysis and Prevention, 2008, No. 1, Vol. 40, pp. 174-181.

[34] Lewis HW. The foundations of fuzzy control, Plenum Publishing Corporation, New York, 1997.

[35] Zadeh LA. Fuzzy Sets, Information and control, California, 1965, No. 3, Vol. 8, pp. 338-353.

[36] Engelbrecht AP. Computational Intelligence: An Introduction. Second Edition, John Wiley & Sons, Ltd, 2007.

[37] Zadeh LA. From computing with words to computing with perceptions, fundamental theory and applications, perceptions, IEEE Journal of Transactions on Circuits and Systems, 1999, No. 1, Vol. 45, pp. 105–119.

[38] Samadzadegan F, Azizi A, Hahn M, Lucas C. Automatic 3D object recognition and reconstruction based on neuro-fuzzy modeling, ISPRS Journal of Photogrammetry and Remote Sensing, 2005, Vol. 59, pp. 255– 277.

[39] Takagi T, Sugeno M. Fuzzy identification of systems and applications to modeling and control, IEEE Trans. On SMC, 1985, No. 7, Vol. 15, pp. 116-132.

[40] Takagi T, Sugeno M. Derivation of fuzzy control rules from human operator's control actions, Proceeding of the IFACSymp on Fuzzy Information, Knowledge Representation and Decision Analysis, 1983, pp. 55-60.

[41] Abraham A. Adaptation of Fuzzy Inference System Using Neural Learning. Fuzzy System Engineering: Theory and Practice, N. Nedjah, Ed. et al. Berlin, Springer-Verlag, Germany, 2005, Chpter 3, pp. 53-83.

[42] Jang JSR. ANFIS: Adaptive-network-based fuzzy inference systems, IEEE Transactions on Systems, Man, and Cybernetics, 1993, No. 3, Vol. 23, pp. 665-685.

[43] Montella A, Aria M, D'Ambrosio A, Mauriello F. Analysis of powered two-wheeler crashes in Italy by classification trees and rules discovery, Accident Analysis and Prevention, 2012, Vol. 49, pp. 58-72.