



## A New Approach to Predict the Severity of Road Accidents with Hybrid MLP ANN and Differential Evolution Algorithm

Zahra Asheghi Dizaji\*, Farhad Soleimanian Gharehchopogh

Department of Computer Engineering, Urmia Branch, Islamic Azad University, Urmia, IRAN.

Receive Date: 10 June 2016; Accepted Date: 22 August 2016, Published Date: 15 September 2016

\*Corresponding Author: zahra\_ashegi@yahoo.com (Z.A. Dizaji)

### Abstract

*In most cases, classification has been proposed as an effective method in decision-making. Although countless methods of classification have been stated by the researchers, none of these methods has the same performance and sufficient accuracy for various issues. Accordingly, a hybrid approach has been proposed for classification in this paper. Also, considering the fact that one of the most common accidents all around the world is road accidents and traffic accidents, which endangers the lives of many people annually, therefore, determining the characteristics affecting the type of the accident is very important. Therefore, the information collected by the transportation and terminal organization in Western Azerbaijan province (Iran) were used as data collection in this paper and the aim of it is the accident classification based on the type of accident (damage, injury, death). For this purpose, a new method has been proposed for classification based on the use of two algorithms of Differential Evolution (DE) algorithm and MLP ANN. In the proposed method, DE algorithm is used for extraction of affecting features of classification, and MLP ANN algorithm is used to classify information based on characteristics determined by DE algorithm. The results of the proposed method have been evaluated with evaluation criteria Recall, Precision, F-Measure, and Kappa which the results show that the proposed method is optimal.*

**Keywords:** DE Algorithm, MLP ANN, Car Accidents, Classification.

### 1. Introduction

Road accidents play a crucial role in limiting economic development and threatening human health. Predicting the type of the accidents is important not only for the governments but also for insurance companies to make decisions about the future. Considering the fact that the number of road accidents is increasing daily in developing countries, this issue is very important in these countries and has challenged many government organs and researchers [1].

Reduction in the number of victims and injuries of the road accidents is an inevitable and important issue in any society. Accordingly, identifying the effective factors in the severity of injuries of road accidents could be an effective step towards

reducing injuries caused by road accidents. In most cases, reduction in the number of accidents is considered as the criterion for prioritizing immunization measures. While the emphasis on the number of accidents causes the ignorance of the impact of immunization measures of the path on the amount of damages caused by accident. Therefore, some methods should be proposed in order to predict the amount of injuries in traffic accidents [2]. The amount of the injuries caused by car accidents is proposed in the three types of damage, injury, and death that accordingly, the amount of injuries caused by road accidents can be categorized in three categories and with no overlap, by data mining and machine learning

algorithms. Algorithms of classification can be applied in various fields of science, and of course each of these algorithms has different functions to the used dataset and the available rules in the dataset. In general, no algorithm has ever been proposed which is able to categorize all the issues with proper accuracy. Accordingly, in this paper a hybrid approach is used considering the importance of determining the amount of injuries in car accidents, and algorithm's function will be discussed in the following. The dataset used in this paper has been gathered from the terminal and transport organization in Western Azerbaijan province (Iran) that this dataset has been adopted based on the real events occurred in this province. In the following, the paper is organized as follows: Section 2, the evaluation of the works done in the field of road accidents and classification of information, Section 3, the basic concepts required for the proposed method, Section 4, the proposed method, Section 5, review and evaluation of the proposed method, and finally, in Section 6 the conclusion is presented.

## 2. Related Works

The importance of reducing the number of accidents has led many researchers to provide solutions to determine the effective factors in accident happenings and moreover, comprehensive safety programs and strategies have been presented by various organs to reduce the number of accidents and fatalities caused by them [3]. In this paper, ANN model is used to identify the factors affecting the amount of the losses and injuries caused by accidents. To develop these strategies, first there should be sufficient knowledge about factors influencing the occurrence of accidents and then the accidents should be modeled based on this knowledge. In recent years, lots of researches have been done in this field and in developed countries and important results were gained [4].

In general, the factors influencing the occurrence of the accidents are categorized into four factors: human, vehicle, road and environment. Since humans have a role in all four of the above-mentioned, the prediction of the number of accidents and their specifications is a complicated work due to the multiplicity of influencing factors and their complex interaction [5]. A variety of methods are presented for modeling the accidents, such as cause and effect methods, linear regression, time series and Markov model [6]. Each of these methods has advantages and disadvantages that can be expressed as follows: in the procedures provided for cause and effect

relationship, a dataset with lots of data is needed to analyze the relationship between influential factors and in methods related to linear regression, the modeling is implemented considering the assumption of "independence of factors of the normal distribution in the prediction process". Also, in the methods of time series the tendencies should be considered stable of the prediction conditions [7]. And finally in Markov model, a probability to move between different conditions should be determined [6]. In line with these methods, new techniques have been proposed for modeling based on machine learning which these methods can be proposed in this way: Marukatat has provided a new method for implementing decision-making support system that is based on adaptive regression trees, that in this way, decision-making support system has the task of analyzing road accidents occurred. In this study, classification, selection and filtering of data is done using rules structure analysis so that the rules structure analysis is done several times according to different parameters and then based on the produced rules in the last stage, the candidate rules are selected [8].

Researchers [9] have used association rules based on particles mass algorithm to specify the influential factors of the accidents which had occurred. In this study, the concept of entropy association is used to compare the importance of the effective factors in accidents and also the T-test model and the Delphi technique are used to test the accuracy of the proposed method and a dataset containing 20 thousand accidents and 56 features is also used. In 2011 Neslihan Karsli et.al. used ANN for modeling driving accidents. In these studies, the best model is selected according to the Akaike Information Criterion (AIC), and has been tested and evaluated using statistical analysis 2 in error conditions and also as a case study in Turkey [10]. In [11] the particles mass algorithm is used to predict road accidents, so that in this study particle mass algorithm is used to determine factors affecting the determination of the amount of injuries caused by road accidents. Also, these researchers have studied the combination of chaos algorithm and particles mass algorithm in another study [12]. In this paper, logistic mapping, Tenet and Lorenz are used as chaos algorithm and also to determine the parameters influencing in the particles mass algorithm.

### 3. Basic Concepts

#### 3.1. MLP ANN

ANNs are very efficient according to the attributes such as high speed, noise immunity, training capability and resistance to the parameters change. The multilayered Perceptron neural network is one of the common techniques in the MLP ANNs, in which the MLP ANN algorithm consists of an input layer, one or more hidden layers and an output layer that the input signals are released in the network and in the path forward layer by layer. This type of ANN, as well as the other methods has activation function in its layers so it has the duty of matching the input signal and output one in each iteration, in terms of weight and neurons bias.

#### 3.2. Differential Evolution Algorithm

DE algorithm was introduced in 1991 by Astorn and Price [13] for the first time. This algorithm, as well as other meta-heuristic algorithms are based on the population so that it uses the distance and direction of the current population to advance the search in the next stages. This algorithm includes the primary initialization and three operators: mutation, intersection and selection. The algorithm is initialized by the Formula (1) and Formula (2).

$$x_i = (x_{i,1}, x_{i,2}, \dots, x_{i,D}) \quad (1)$$

$$\begin{aligned} x_{i,k} &= x_k^{min} + rand(0,1) \\ &* (x_k^{max} - x_k^{min}) \quad \text{with } i \in [1, NP], k \in [1, D] \end{aligned} \quad (2)$$

Choosing random numbers of  $x_{i,k}$  from the problem domain is done using the Formula (2). In Formula (2), D equals to the dimensions of the solutions and the variable Np is the number of initial population. Rand (0, 1) Function produces random numbers (with uniform distribution) in the interval (0,1). It is also clear that if the Formula (2) is used, obtained values for  $x_{i,k}$  will be in the range  $[x_i^{max}, x_i^{min}]$  and the position vector of each of the solutions will be a potential answer for the optimization problem.

The main difference between DE and Genetic Algorithm is the mutation operator. In this algorithm, the Formula (3) is used for mutation operator.

$$v_{i,G+1} = x_{r1,G} + F \cdot (x_{r2,G} - x_{r3,G}) \quad (3)$$

In Formula (3), r1, r2, r3 are three unequal random numbers in the range  $[Np, 1]$ , G is the number of produced generations and coefficient F is a positive and real constant number which is often considered 0, 5.

Integration operator leads to increase in the diversity of the population. This operator is similar to the integration operator in genetic algorithm that uses Formula (4) to perform integration.

$$u_{ij,G+1} = \begin{cases} v_{ij,G+1} & \text{if } (r_j \leq CR) \text{ or } j = j_{rand} \\ x_{ij,G+1} & \text{otherwise} \end{cases} \quad (4)$$

In Formula (4), CR parameter in the range  $[0,1]$  and  $r_j$  parameter in the range  $[0,1]$  are randomly produced. And also the amount of j equals 1,2, ..., D.

In order to select the vectors with high fitness, vectors which were produced by the operators of mutation and integration are compared with each other, and whichever is more fit will be transferred to the next generation.

### 4. Proposed Method

With the ever-increasing number of road accidents and complexity of relations between factors in occurrence of accidents, it is necessary to provide new ways to predict accidents. Meanwhile, given that machine-based learning methods are based on population and repetition, they are able to provide the desired results. Accordingly, in this paper, a hybrid method based on MLP ANN algorithm and DE algorithm is proposed to determine the affecting factors in the amount of damages and injuries caused by the accidents. In fact, in the proposed method, the used dataset is classified into three categories of damage, injury and death based on the amount of injuries resulting from accidents. In the proposed method, after reading the dataset, the data are presented to the DE algorithm at the first level and this algorithm selects the factors with more importance from the effective factors listed (people, vehicles, roads and environment) in the dataset, and in order to select the most important factors after initialization and crossover operations and mutation from the K algorithm, the nearest neighbor is selected as the cost function to calculate the accuracy size of the factors. At this stage, after determining the extent of accuracy, the lowest value is selected as the best solution. After finishing working with DE algorithm, the new dataset is produced according

to effective factors selected and then the new dataset is divided into two training and testing datasets with the ratio of 80 to 20 percent and it is given to the ANN algorithm as an input.

In the ANN algorithm, three layers (input, hidden, output) and the error propagation algorithm and logistic activation function have been used as

activator function between neurons. ANN is responsible for the classification of dataset produced by DE algorithm, which after the classification, the data classified by evaluation criteria are evaluated and then the necessary charts and tables are produced based on these criteria.

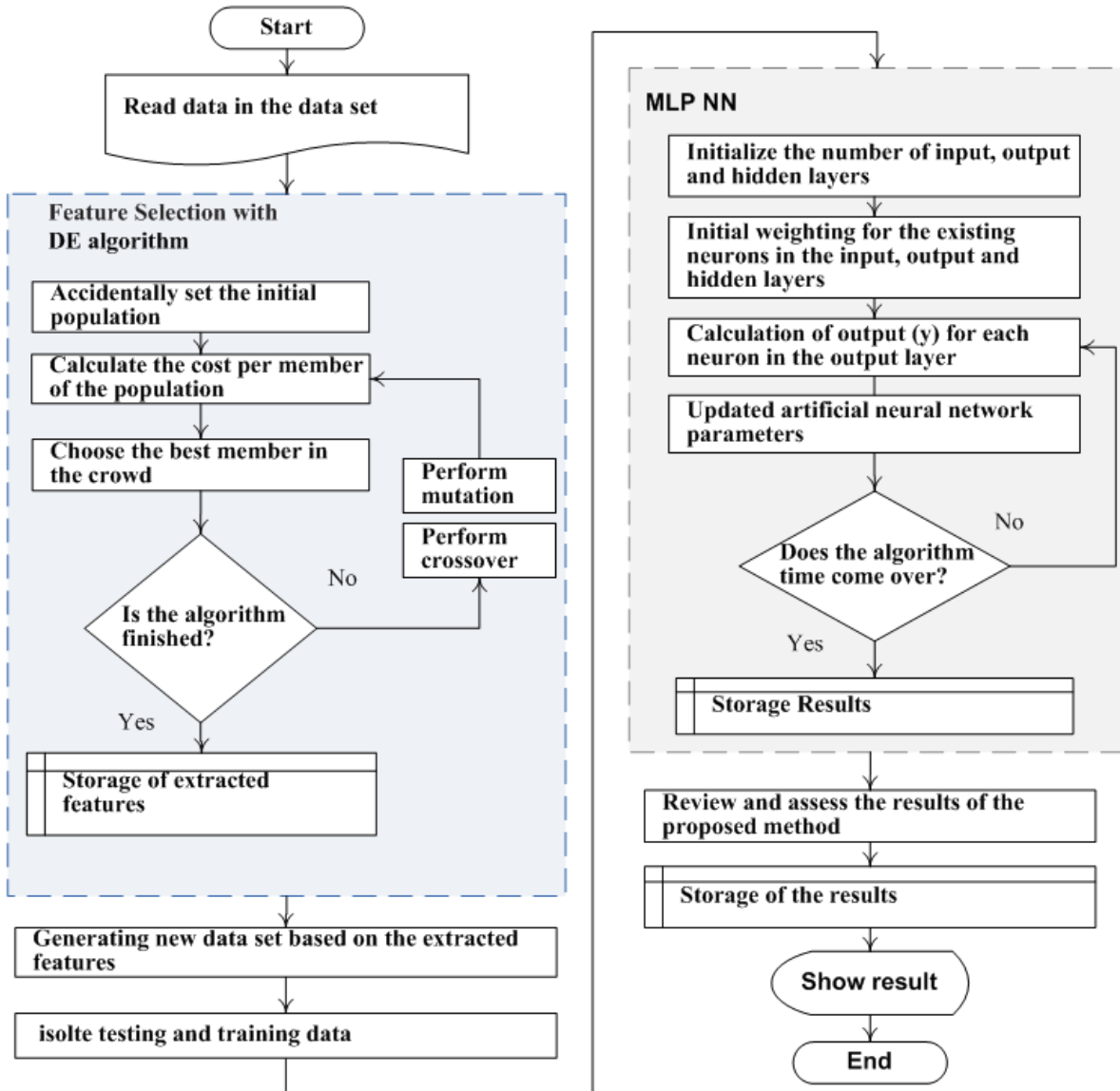


Figure 1. Flowchart of Proposed Method

**5. Result and Discussion**

With presentation of new methods by researchers, different evaluation criteria should be used in order to evaluate the quality of the proposed method, which in this paper, Precision evaluation criteria, recall and F-Measure are used to evaluate the proposed method [14, 15]. In the Formula (5),

Formula (6) and Formula (7) these three evaluation criteria are expressed.

$$Recall = \frac{No. Of relevant images retrieved}{Total no. of relevant image in the collection} \tag{5}$$

$$Precision = \frac{No. of relevant images retrieved}{Total no. of images retrieved} \tag{6}$$

$$F_{measure} = 2 * \frac{Percision * Recall}{Percision + Recall} \tag{7}$$

As mentioned at the beginning of this paper, the dataset collected by the terminal and transportation organization of the Western Azerbaijan province (Iran) is used in this paper. The results of the proposed method are shown in Table 1 and Table 2.

**Table 1. Results obtained from the training data set**

Method	Evaluation Criteria	Result
MLP	Recall	0.980
	Precision	0.8483
	F-Measure	0.9179
	Kappa	0.8357
	Categorized correctly	547
	Not categorized correctly	32
Proposed Method	Recall	0.982
	Precision	0.8593
	F-Measure	0.9243
	Kappa	0.8822
	Categorized correctly	550
	Not categorized correctly	29

In Table 1, the proposed method is evaluated on the training dataset. According to the results, 550 of the accidents are classified correctly based on the injuries and only 29 cases of data have not been categorized correctly. And oppositely, the MLP algorithm correctly classified 547 cases of accidents based on the injuries and only 32 cases of data are not categorized correctly.

**Table 2. Results Obtained from Testing Dataset**

Method	Evaluation Criteria	Result
MLP	Recall	0.99
	Precision	0.8947
	F-Measure	0.9444
	Kappa	0.9161
	Categorized correctly	139
	Not categorized correctly	6
Proposed Method	Recall	0.992
	Precision	0.9333
	F-Measure	0.9655
	Kappa	0.93
	Categorized correctly	141
	Not categorized correctly	4

In Table 2, the proposed method is evaluated on the test dataset. According to the results, 141 of the accidents are classified correctly based on the injuries and only 4 cases of data have not been categorized correctly. And oppositely, the MLP algorithm correctly classified 139 cases of accidents based on the injuries and only 6 cases of data are not categorized correctly.

**6. Conclusion**

Many factors affect the occurrence of road accidents and these factors have complicated relations in a way that manually finding these

relations is very difficult. Accordingly, in this paper, a new hybrid method is proposed based on MLP ANN algorithm and DE algorithm. Initially, factors affecting the determination of the amount of injuries caused by accidents are extracted by DE algorithm in the suggested method and then the new dataset produced based on the extracted features is presented as the input to the MLP ANN algorithm. And finally, the data classified by MLP ANN algorithm (damage, injury and death) are evaluated by the defined evaluation criteria. Based on the results of the suggested method, it can be said that the combination of MLP ANN and DE algorithms, the performance of MLP ANN algorithm is improved.

**References**

- [1] M.Penden, R.Scurfield, D.Sleet, D.Mohan, AA.Hyder, E.Jarawan, C. Mathers, World Report on Road Traffic Injury Prevention, World Health Organization, Pages. 244, 2004.
- [2] D.Shinar, Safety and Mobility of Vulnerable Road Users: Pedestrians, Bicyclists, and Motorcyclists, [Editorial], Accid Anal Prev, Vol.44, No.1, pp.10-20, 2012.
- [3] D.Delen, R.Sharda, and M.Bessonov, Identifying Significant Predictors of Injury Severity in Traffic Accidents Using a Series of Artificial Neural Networks, Accident Analysis And Prevention, Vol.38, No.3, pp.434-444,2006.
- [4] S. M.Rifaat, R.Tay and A.De Barros, Effect of Street Pattern on the Severity of Crashes Involving Vulnerable Road Users, Accident Analysis and Prevention, Vol.43, No.1, pp.276- 283, 2011.
- [5] C.Oh, J. S.Oh, S. G.Ritchie and M .Chang, Real-time Estimation of Freeway Accident Likelihood, Annual Meeting CD-Rom 80th Annual Meeting of The Transportation Research Board, Washington, DC, pp.1–16, January 7-11, 2001.
- [6] H. Li-Chang, Using Improved Grey Forecasting Models to Forecast the Output of Opto-Electronics Industry, Expert Systems with Applications, Vol. 38, pp.13879–13885, 2011.
- [7] K. Erdal, U. Baris, and K. Okyay, Grey System Theory Based Models in Time Series Prediction, Expert Systems with Applications, Vol. 37, pp.1784–1789, 2010.

- [8] R. Marukatat, Structure-Based Rule Selection Frame Work for Association Rule Mining of Traic Accident Data, Computational Intelligence and Security, Vol.4456, pp.231–239, 2007.
- [9] J. Xi, Z. Gao, S. Niu, T. Ding, and G. Nin, A Hybrid Algorithm of Traffic Accident Data Mining on Cause Analysis, Mathematical Problems in Engineering, Vol.2013, pp.1-8, 2013.
- [10] H.F. Bayata, F. Hattatoglu, and N. Karsli Modeling of Monthly Traffic Accidents with the Artificial Neural Network Method, International Journal of the Physical Sciences, Vol. 6, No.2, pp. 244-254, 2011
- [11] F.S. Gharehchopogh, Z.A. Dizaji, and Z.Aghighi, Evaluation of Particle Swarm Optimization Algorithm in Prediction of the Car Accidents on the Roads: A Case Study, International Journal on Computational Sciences & Applications (IJCSA), Vol.3, No.4, pp.1-12, 2013.
- [12] F.S. Gharehchopogh, Z.A. Dizaji, A New Chaos Agent Based Approach in Prediction of the Road Accidents with Hybrid of PSO Optimization and Chaos Optimization Algorithms: A Case Study, International Journal of Academic Research Part A, Vol.6, No.2, pp.108-115, 2014.
- [13] R. Storn and K. Price, Minimizing the Real Functions of the ICEC'96 Contest by Differential Evolution, International Conference on Evolutionary Computation, Nagoya, Japan, pp. 842-844, 1996.
- [14] D.M. Powers, Evaluation: from Precision, Recall and F-measure to ROC, In formedness, Markedness and Correlation, Journal of Machine Learning Technologies Vol.2, No.1, pp.37–63, 2011.
- [15] F.S. Gharehchopogh, I.Amini, B.Zeberdast, A Three-Layer Architecture based Approach for Data Access Layer in the Information Systems Production, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Vol: 2, No: 2, pp: 761-766, February 2013.