

Article

Evaluation of Contributing Factors Affecting Number of Vehicles Involved in Crashes Using Machine Learning Techniques in Rural Roads of Cosenza, Italy

Giuseppe Guido ¹, Sina Shaffiee Haghshenas ¹, Sami Shaffiee Haghshenas ¹, Alessandro Vitale ¹, Vittorio Astarita ^{1,*}, Yongjin Park ^{2,*}, and Zong Woo Geem ³

¹ Department of Civil Engineering, University of Calabria, Via Bucci, 87036 Rende, Italy; Giuseppe.guido@unical.it (G.G.); Sina.shaffieehaghshenas@unical.it (S.S.H.); shfsma87c29z224h@studenti.unical.it (S.S.H.); Alessandro.vitale@unical.it (A.V.); vittorio.astarita@unical.it (V.A.)

² Department of Transportation Engineering, Keimyung University, Daegu 42601, Korea; ypark@kmu.ac.kr (Y.P.)

³ College of IT Convergence, Gachon University, Seongnam 13120, Korea; geem@gachon.ac.kr (Z.W.G.)

* Correspondence: vittorio.astarita@unical.it (V.A.); ypark@kmu.ac.kr (Y.P.); Tel.: +39-098-449-6780 (V.A.)

Abstract: Evaluation of road safety is a critical issue having to be conducted for successful safety management in road transport systems, whereas safety management is considered in road transportation systems as a challenging task according to the dynamic of this issue and the presence of a large number of effective parameters on road safety. Therefore, evaluation and analysis of important contributing factors affecting the number of crashes play a key role in increasing the efficiency of road safety. For this purpose, in this research work, two machine learning algorithms including the group method of data handling (GMDH)-type neural network and a combination of support vector machine (SVM) and the grasshopper optimization algorithm (GOA) are employed for evaluating the number of vehicles involved in the accident based on the seven factors affecting transport safety including the Daylight (DL), Weekday (W), Type of accident (TA), Location (L), Speed limit (SL), Average speed (AS) and Annual average daily traffic (AADT) of rural roads of Cosenza in southern Italy. In this study, 564 data sets of rural areas were investigated and relevant effective parameters were measured. In the next stage, several models were developed to investigate the parameters affecting the safety management of road transportation for rural areas. The results obtained demonstrated that "Average speed" has the highest level and "Weekday" has the lowest level of importance in the investigated rural area. Finally, although the results of both algorithms were the same, the GOA-SVM model showed a better degree of accuracy and robustness than the GMDH model.

Keywords: Road safety; Safety management; Road transportation; GMDH; GOA-SVM; Machine learning.

1. Introduction

Road transport is one of the oldest ways of transporting goods and individuals. With the increase in population and the development of towns, the expansion of road transport has become an inevitable issue that plays a significant role in increasing and improving economic and social development. Numerous issues affect the quality and quantity of road transport, and safety management is one of the most significant subjects. Therefore, identifying, studying, and evaluating the contributing factors affecting road safety is an inevitable thing to increase the level of safety management. Several parameters affect road safety and valuable studies have been conducted to investigate road safety, such as Driver Behavior [1-6], Age of Driver and Vehicle [7-12], Weather Conditions [13-17], Road

Geometry [18-22], Lighting Conditions [23-28]. Zoe et al. (2018) carried out an efficient analysis of the expansion trend on road safety researches between 2000 and 2018. Their results indicated that road safety research works focused on five major areas including accident frequency data investigation, driver behavior questionnaire, safety in numbers for walkers and bicyclists, injury and prevention of road traffic, and driving speed and road accidents [29].

Moreover, other comprehensive studies have been done to examine the role of other factors on road safety. Gichaga (2017) reviewed the historical and cultural background involving road development and road safety features in Kenya. Based on his results, he made some recommendations for improvements in aspects of road safety [30]. Elvik et al., (2019) conducted a review of the relationship among speed and road safety. They supported two mathematical models. Their results showed that the speed of individual drivers has a similar relationship with safety as the mean speed of traffic [31]. In another study, the effect of air quality on road safety was evaluated by Sager., (2019). He investigated the impact of increased air pollution on the amount of road traffic crashes. Then, he found out that there is a relationship among the number of accident and the amount of PM2.5 [32].

On the other hand, there are several techniques to evaluate the parameters affecting road safety. Traditionally, classical before-after research works, statistical modeling, and personal judgment-based approaches are applied for chronological data [33]. Zheng et al., (2018) investigated to propose a bivariate extreme value model. Their proposed model could decrease the uncertainty in crash estimates [34]. Also, according to a bivariate extreme value theory (EVT) framework, Wang et al., (2019) introduced an accident forecast method. Their results prove that the proposed crash prediction method can provide very promising results compared with univariate models [35]. In another research work, Zheng et al., (2020) carried out a comprehensive review of researches, based on the use of traffic conflicts in the road safety analysis. Then, they discussed conceptual and methodological matters related to traffic conflict modeling. Based on their results, it has been determined that although suitable research studies have been done on this issue, the need for more researches is necessary [36]. An overview of Road traffic accidents was conducted in the Eastern Province, KSA, from 2009 to 2016 by Jamal et al., (2020). They developed logistic regression models to forecast crash severity. Finally, they recommended some suggestions to prevent road accidents [37].

Due to the uncertainty and unforeseen problems that there are like most parameters affecting road safety and also the high ability of machine learning algorithms, recently the use of these methods in combination with classical methods or alone in many studies have been widely used [38-46]. Xu et al., (2018) evaluated the effect of road lighting on road safety by an artificial neural network. The results clearly showed that road lighting had a great contribution to road safety levels [47]. In another study, two types of artificial intelligence methods were applied to estimate severity crashes by Amiri et al., (2020). They used the crash information of California in 2012. Their results indicated that both artificial intelligence models were reliable for predicting severity crashes and also, the light condition played an important role in the level of severity in comparison with the other contributing factors [48]. Guido et al., (2020) investigated some potential factors for road safety using two artificial intelligence methods. Their results indicated that the PSO algorithm had a superior function in comparison with the GA algorithm for evaluating factors in road accidents [49]. Shiran et al., (2021) carried out crash severity analysis by applying data mining approaches and multinomial logistic regression. They used an accident dataset of State Highways in California, USA. They found out that the C5.0 model provides higher performance capacity in evaluating crash severity analysis compared to other models [50].

Studies of the past literature show that although valuable studies have been done, but based on the increasing expansion of road transport, more research is needed to increase the quality of road transport. Therefore, in this study, two machine learning algorithms including the group method of data handling (GMDH)-type neural network and a combination of support vector machine (SVM) and the grasshopper optimization

algorithm (GOA) are applied. Then, the obtained results of two algorithms are compared based on performance indicators to determine the performances of models regarding the conditions and characteristics of a case study. Finally, a sensitivity analysis is conducted for prioritizing the contributing factors affecting the number of crashes in a rural area.

2. Site description and accident monitoring

To evaluate the performance and the usefulness of the proposed approaches for prioritizing the contributing factors affecting the number of crashes in a rural area, a sample of 564 accident data was acquired from 2017 to 2018 on the rural road network of Cosenza province has been analyzed (Fig. 1).

The accident data have been acquired from the database of the Automobile Club Italia (ACI) which collaborates with the National Institute of Statistics (ISTAT) in collecting road accidents data in Italy. This database provides a lot of information about accidents, such as date and place of accident, road category, pavement conditions, weather conditions, dynamics of the crash, the type of vehicles involved, the causes of the accident, and the consequences for the people involved (injuries or deaths). The information does not include Property Damage Only (PDO) events under existing Italian legislation, which defines road accidents as accidents only when they cause at least one injury [51].

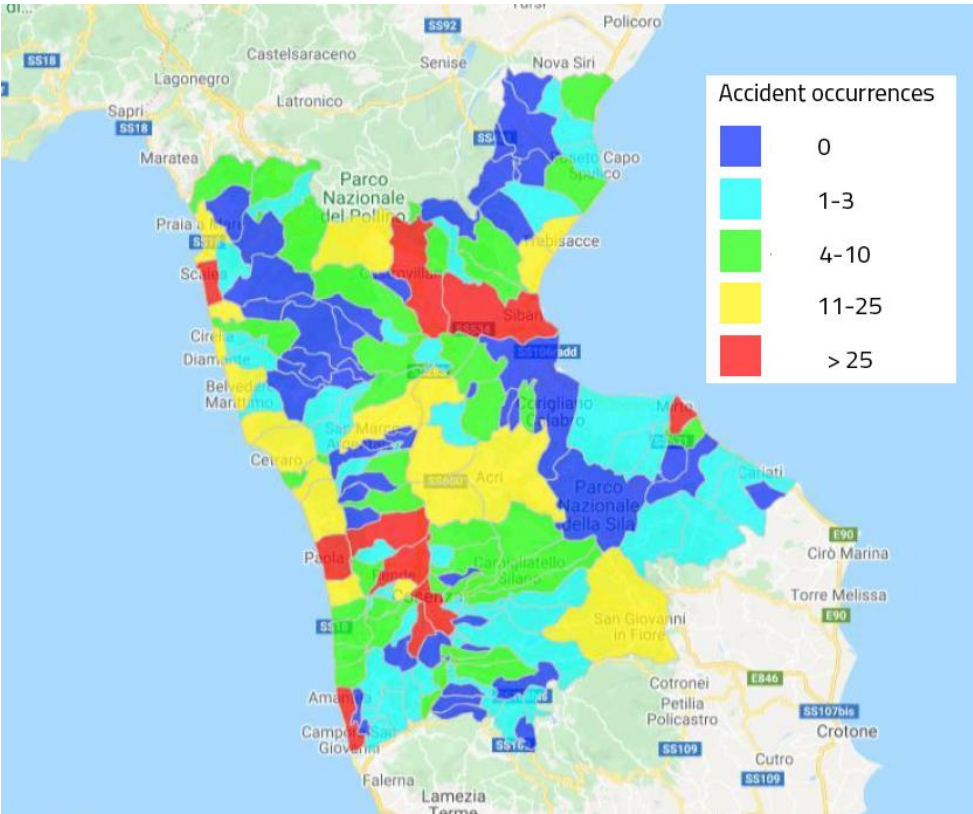


Figure 1. Rural roads accident map in the province of Cosenza for years 2017 and 2018 (source: Regional Center for the Collection of Data on Road Accidents in Calabria – CRISC).

According to ISTAT, accidents are classified as fatal or with injuries; therefore, it is not possible to distinguish the injured according to the level of severity. Based on the previous comment, in this paper, the measure of the crash severity, as referred to in section 4, is expressed as a dichotomous variable (0,1):

- 0 when the accident accounts for one injury;
- 1 when the accident accounts for more than one injury or a fatal accident.

To ensure a more detailed analysis, other information on roads and traffic flows was considered, such as speed limits, average speed, and Average Annual Daily Traffic

(AADT). The speed limits are regulated by the “Nuovo Codice della Strada” [51-52], and their values have been acquired from the database of Azienda Nazionale Autonoma delle Strade (ANAS). The average speed was obtained through the available data of the historical traffic statistics of TomTom (TomTom Move) and Octo Telematics (Octo IoT Cloud) referring to the road sections with observed accidents. The AADT was acquired from the PANAMA system of ANAS (ANAS Platform for Monitoring and Analyzing).

This study considers seven independent variables (i.e. the factors affecting the number and the severity of the accidents), which were selected based on all the available data for the case study. These variables include four qualitative variables (Daylight (DL), Weekday (W), Type of accident (TA), Location (L)) and three quantitative variables (Speed limit (SL), Average speed (AS), Annual Average Daily Traffic (AADT)).

Table 1 shows the above-mentioned variables, which are classified into various categories (codes) with their statements. The group method of variables into different categories depends on their characteristics and the amount of observations involved in each study.

Table 1. Independent variables (qualitative and quantitative).

Data field type	Variable	Code/Unit	Description
Traffic flow characteristics	AADT (veh/day)	1	<5,000
		2	5,000 – 9,999
		3	10,000 – 14,999
		4	>14,999
	Avg Speed (km/h)	Not coded	Min 28 Max 122 Avg 91.43
Road environment	Location	0	Non intersection
		1	Intersection
	Speed Limit (km/h)	1	50
		2	70
		3	90
		4	110
Environment characteristics	DayLight	5	130
		0	Daylight
	Weekday	1	Nighttime
		0	Weekend or Holiday
		1	Weekday
Accident characteristic	Accident Type	1	Collision with vehicle
		2	Collision with pedestrian
		3	Collision with obstacle
		4	Other

3. Methodology

The study of road safety is one of the inseparable issues of transportation engineering, which is usually defined by the absence of accidents and casualties. Lack of attention, to road safety, could impose irreparable financial and physical damages, therefore, it is necessary to have a deep understanding of road safety and to know all the effective components, and estimate the impact of each of them on this issue. By studying the literature reveals that the majority of investigations in the field of road safety are based on logit models or regression models with artificial neural networks. On the one hand, due to the complexity and uncertainty in the factors affecting road safety, and on the other hand, the

ability of machine learning algorithms in the face of unpredicted and uncertain issues has led to the successful application of machine learning methods was used to road safety in recent years. [54-60].

Accordingly, the main aim of this study is to investigate the factors affecting road safety in rural areas by using two machine training approaches, namely the GMDH model and the hybrid GOA-SVM model. For this purpose, the GMDH model is developed to achieve the best binary classification model by determining the best control parameters of GMDH. Also, the SVM was hybridized with the grasshopper optimization algorithm as a suitable evolutionary algorithm and was developed to optimize its three parameters of SVM. Finally, a sensitivity analysis is performed to evaluate and rank factors affecting the number of accidents in rural areas. More discussions regarding machine learning models will be given in the next sections.

3.1. Group method of data handling-type neural network

In today's science, ANNs (as one of the branches of artificial intelligence) play a very valuable role in the development of new technologies. Therefore, the use of ANNs to solve many complex problems in various fields of science is inevitable [61-66]. Group method of data handling (GMDH) is one of the artificial neural networks that was introduced by Ivakhnenko (1971). GMDH was applied for computer-based mathematical modeling in complex systems and also, it has been used successfully for data mining, optimization, and pattern recognition problems [67]. The process of GMDH is similar to a type of self-organizing network [68-69]. The mapping among the input and output variables in GMDH neural network is a nonlinear function that Polynomial Neural Network (PNN) is one of the basic algorithms used to construct GMDH models. In GMDH modeling, input data are entered into the initial layer and after preparation are considered as input for the second layer and this process continues until the algorithm converges and stops. Finally, in the convergence process of the algorithm, if the results in the layer (n+1) are better than the layer (n), then the algorithm converges. Eqs 1 and 2 indicate the relationship among the approximate function (\hat{f}) with the multi-input and single-output (\hat{y}) dataset and the least possible error between actual and predicted values [70-71].

$$\hat{y} = \hat{f}(x_{i1}, x_{i2}, x_{i3}, \dots, x_{im}) \quad (1)$$

$$i = (1, 2, 3, \dots, n)$$

$$\sum_{i=1}^M \left[\hat{f}(x_{i1}, x_{i2}, x_{i3}, \dots, x_{im}) - y_i \right]^2 \Rightarrow \text{Min} \quad (2)$$

Eq 3 indicates a relationship between a single output (y) and a multi-input (Input vector, $X = (x_1, x_2, x_3, \dots, x_m)$) according to the Kolmogorov-Gabor polynomial [72-73].

$$y = a_0 + \sum_{i=1}^m a_i x_i + \sum_{i=1}^m \sum_{j=1}^m a_{ij} x_i x_j + \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m a_{ijk} x_i x_j x_k + \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m \sum_{l=1}^m a_{ijkl} x_i x_j x_k x_l, \dots \quad (3)$$

Where $a_i, a_{ij}, a_{ijk}, a_{ijkl}, \dots$ are considered coefficients of polynomial and m is the amount of data. Furthermore, Eq 3 can be considered as a quadratic polynomial for 2 Inputs according to Eq 4 [74].

$$\hat{y} = G(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i^2 + a_4 x_j^2 + a_5 x_i x_j \quad (4)$$

The total error (E) is considered by minimizing the difference among the actual output (y) and predicted output ($\hat{y} = G(x_i, x_j)$) for each pair of input variables x_i and x_j based on Eq 5. Moreover, the coefficients of each quadratic function are optimized [75].

$$E = \frac{\sum_{i=1}^M (y_i - G_i(x_i, x_j))^2}{M} \Rightarrow \text{Min} \quad (5)$$

Out of a total of n input variables, all alternatives for two independent variables are provided in the elementary form of the GMDH algorithm for providing the regression polynomial with the form of Eq 4 [76]. Therefore, $\binom{n}{2} = \frac{n(n-1)}{2}$ neurons will be made in the primary layer of the feedforward neural network from the observations $\{(y_i, x_i, x_{iq}); (i=1, 2, 3, \dots, M)\}$ for various $(p, q \in \{1, 2, 3, \dots, n\})$. The matrix form of Eq 4 can be considered to indicate the main form of the GMDH, based on Eq 6 [75].

$$Y = Aa \quad (6)$$

$Y = (y_1, y_2, y_3, \dots, y_m)$ and $a = (a_1, a_2, a_3, a_4, a_5)$ are considered the observed output vector and the vector coefficient of the quadratic polynomial, respectively. A is computed based on Eq 7 [77].

$$A = \begin{bmatrix} 1 & x_{1p} & x_{1q} & x_{1p}x_{1q} & x_{1p}^2 & x_{1q}^2 \\ 1 & x_{1p} & x_{1q} & x_{1p}x_{1q} & x_{1p}^2 & x_{1q}^2 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_{Mp} & x_{Mq} & x_{Mp}x_{Mq} & x_{Mp}^2 & x_{Mq}^2 \end{bmatrix} \quad (7)$$

Finally, using the least-squares process from multiple regression analysis, a normal equation is achieved based on Eq 8, which calculates the vector of the best coefficients for Eq 4 [78].

$$a = (A^T A)^{-1} A^T Y \quad (8)$$

3.2. Support vector machine

Support vector machine (SVM) is an effective machine learning method which was introduced by Cortes & Vapnik (1995) [79]. SVMs are a kind of supervised learning algorithms which are used in a wide range of modelings, such as regression and classification. The SVM presents a linear two-class classifier and it aims to maximize the margin amongst two classes, so that a classification hyperplane is made in the central of the maximum margin. It provides many hyperplanes, while the goal of the support vector machine is to find the best hyperplane in n-dimensional space. Two labels are considered for this

classification that label +1 is considered for cases that are above the hyperplane and the label -1 is belongs to cases that are under the hyperplane. Eq 9 shows a group of the sample set that is used in classification learning data [80-81].

$$S = \left\{ (x_i, y_i)_{i=1}^n \mid x_i \in R^N, y_i \in \{-1, 1\}, i = 1, 2, \dots, l \right\} \quad (9)$$

Where y_i is the target variable for the observed i th sample (The sample category). It is also assumed that x_i presents the i th sample data. After the formation of hyperplanes, one of the hyperplanes has the highest margin, which is called the optimal hyperplane. This optimal hyperplane is determined by the existing support vectors and constraints. Eqs 10 and 11 indicate the constraints [82-83].

$$\text{Min} \frac{1}{2} \|w\|^2 \quad (10)$$

$$\text{s.t. } y_i (wx_i + b) \geq 1 \quad (11)$$

Where w and b are the weight vector and the bias vector, correspondingly. Then, considering an error coefficient, the constraints are rewritten and corrected according to Equations 12 and 13. This error coefficient is intended to ensure a more accurate classification [84].

$$\text{Min} \frac{1}{2} \|w\|^2 + c \sum_{i=1}^n \varepsilon_i \quad (\varepsilon_i \geq 0) \quad (12)$$

$$\text{s.t. } \begin{cases} y_i (wx_i + b) \geq 1 - \varepsilon_i \\ c \geq 0 \end{cases} \quad (i = 1, 2, 3, \dots, n) \quad (13)$$

Where c is the penalty coefficient. Then, using the Lagrange method, classification problems of SVM are considered as the following dual optimization problem based on Eq 14 [85].

$$\begin{cases} W(a) = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i,j=1}^n a_i a_j y_i y_j K(x_i, x_j) \\ \text{s.t. } \sum_{i=1}^n a_i y_i \quad (0 \leq a_i \leq c; i = 1, 2, 3, \dots, n) \end{cases} \quad (14)$$

Where K is a mathematical function that is called the kernel function. There are different types of kernel functions including the linear (LIN), the radial basis function (RBF), the polynomial (POL), and their relationships are shown in Table 2. Gamma (γ) and d are necessary to define the kernel types that Gamma (γ) is used for RBF and POL and " d " represents the term of polynomial degree only for the POL kernel function [86-87]. The most important role of the kernel function is to take the dataset as input and convert it into the required form. Knowledge of the use of various kernel functions in related situations can affect the quality of the category.

Table 2. Equations of different kernel functions.

No	Type of Kernel function	Equations
----	-------------------------	-----------

1	Linear (LIN)	$G(x_i, x_j) = x_i^t x_j$
2	Radial basis function (RBF)	$G(x_i, x_j) = \exp(-\gamma \ x_i - x_j\ ^2)$
3	Polynomial (POL)	$G(x_i, x_j) = (-\gamma x_i^t x_j + 1)^d$

3.3. Grasshopper optimization algorithm

In recent years, meta-heuristic algorithms have played a very important role in dealing with complex and uncertain problems [88-93]. These algorithms have made significant progress in both academia and industry. Grasshopper optimization algorithm (GOA) is one of the newest meta-heuristic algorithms which was presented by Saremi et al., (2017) [94]. This algorithm is according to swarm intelligence and population-based which was inspired by the group behavior of grasshoppers. Grasshoppers are usually seen individually or in groups. One of the most important features of the group of grasshoppers is their type of movement, which has a slow movement with small steps. In the locust algorithm, two sections are defined for search, including exploration and exploitation. Search agents are persuaded to move suddenly in exploration, whereas in exploitation, they want to move locally. To simulate the swarming behavior of locusts, a mathematical model was introduced according to Eq 15 [94-95].

$$X_i = S_i + G_i + A_i \quad (15)$$

Where X_i represents the position of the i-th grasshopper. S_i , G_i and A_i are the social interaction, the gravity force on the i-th grasshopper, and the wind advection, respectively. Then, to show the random behavior of grasshopper in Eq 15, the random factors are used including r_1 , r_2 , and r_3 that are random values within [0,1]. Eq 15 is rewritten according to Eq 16 [94], [96].

$$X_i = r_1 S_i + r_2 G_i + r_3 A_i \quad (16)$$

The gravity force on the i-th grasshopper (S_i) is presented based on Eq 17 [94], [96].

$$S_i = \sum_{\substack{j=1 \\ j \neq i}}^N s(d_{ij}) \hat{d}_{ij} \quad (17)$$

Where d_{ij} and \hat{d}_{ij} present the distance among i-th and j-th grasshopper and a unit vector from the i th grasshopper to the j th grasshopper that is computed as

$$d_{ij} = |x_j - x_i| \text{ and } \hat{d}_{ij} = \frac{x_j - x_i}{d_{ij}}, \text{ respectively. } s \text{ is a function to describe the strength of}$$

social forces based on Eq 18. The motion of the grasshopper is affected by the repulsion and attraction factor between them, which is defined by s in the mathematical model of GOA based on Fig. 2. According to Fig 1, there is a comfort zone (comfortable distance) in which neither the attraction nor the repulsion action takes place between two grasshoppers [94].

$$s(r) = f e^{\frac{-r}{l}} - e^{-r} \quad (18)$$

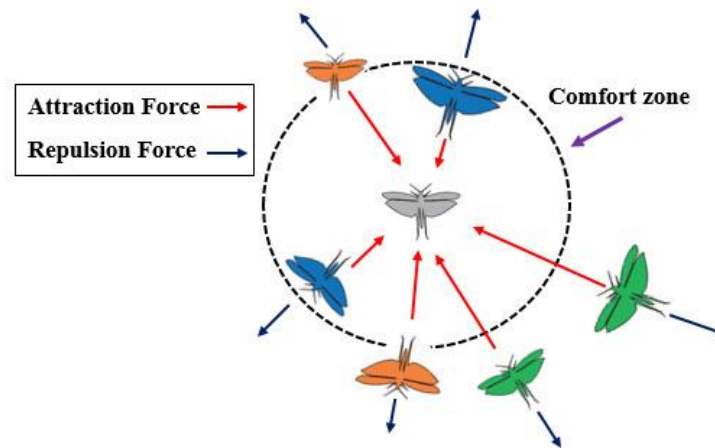


Figure 2. Attraction force, repulsion force, and comfort zone in primary corrective patterns among individuals in a swarm of grasshoppers.

Where f and l are the intensity of attraction and the attractive length scale, correspondingly. Changes in the amount of f and l indicate a change in the behavior of the grasshopper, so by changing them, the number of S changes, and the final results can change. Eqs 19 and 20 show the gravity force on the i -th grasshopper and the wind advection, respectively [94].

$$G_i = -g\hat{e}_g \quad (19)$$

$$A_i = u\hat{e}_w \quad (20)$$

Where \hat{e}_g is a unity vector towards the center of earth and g is the constant of gravitation. Also, u and \hat{e}_w present a constant drift and a unity vector in the direction of the wind, correspondingly. It should be noted that nymph grasshoppers do not have wings, so their movements are strongly influenced by the wind. By inserting the values of each of these definitions, the mathematical model of the algorithm expands based on Eq 21 [94], [97].

$$X_i = \sum_{\substack{j=1 \\ j \neq i}}^N s(|x_j - x_i|) \frac{x_j - x_i}{d_{ij}} - g\hat{e}_g + u\hat{e}_w \quad (21)$$

It should be noted that if the locust population reaches the comfort zone quickly, the swarm will not converge to a specific point, and Eq 21 is not able to solve optimization problems directly. Therefore, the mathematical model presented in Eq 21 considers an upper bound and lower bound as well as two coefficients to balance the motion between the comfort, gravity, and repulsion zones, and Eq 21 is modified as Eq 22 [94], [98].

$$X_i^d = c \left(\sum_{\substack{j=1 \\ j \neq i}}^N c \frac{ub_d - lb_d}{2} s(|x_j - x_i|) \frac{x_j - x_i}{d_{ij}} - g\hat{e}_g + u\hat{e}_w \right) + \hat{T}_d \quad (22)$$

Where ub_d and lb_d represent upper bound and lower bound in the D th dimension, respectively. \hat{T}_d presents the value of the D th dimension in the target and c introduces a reducing coefficient to shrink the comfort zone, repulsion zone, and attraction zone that is shown in Eq 23 [94], [99].

$$c = c_{\max} - l \frac{c_{\max} - c_{\min}}{L} \quad (23)$$

Where c_{max} and c_{min} are the maximum value and the minimum value, correspondingly. Also, l demonstrates the current iteration, and L shows the maximum number of iterations. Parameter c is needed to be reduced proportionally to the amount of iterations for balancing exploration and exploitation. For more information and explanations regarding the grasshopper optimization algorithm refer to Saremi et al. (2017) [94].

4. Results and discussion

As mentioned earlier, for the binary classification modeling, the two machine learning algorithms namely GMDH and the hybrid GOA-SVM were used and developed, and the number of vehicles involved in the accident was evaluated in this study. For this purpose, a valuable dataset was collected and, as described in section 2, the seven factors affecting the number of vehicles involved in the crashes including the Daylight (DL), Weekday (W), Type of accident (TA), Location (L), Speed limit (SL), Average speed (AS) and Annual average daily traffic (AADT) of rural areas of Cosenza in southern Italy were considered. In this study, all data have been classified into two classes. To check the number of vehicles involved in the accident, the first class with the label "0" was considered for cases where at most one car was involved in an accident. And the second class was labeled "1" for cases in which at least two or more vehicles were involved in a crash. This classification was based on the assumption that the main criterion for class separation is to consider the minimum number of vehicles involved in an accident. This was done by developing and constructing the best classification model to determine the correct classes with the highest possible accuracy by determining a mapping between the input and output data.

By developing models, the best models of each method are determined and then the results obtained are compared. Finally, by performing a sensitivity analysis, the importance of the effect of each of the factors is determined. It is necessary to mention that in binary classification modeling, the use of accuracy and error of confusion matrix is considered as the most practical performance indicators. Therefore, to compare the performance evaluation of the models, the confusion matrix is used according to Fig. 3 and Eqs 24 and 25. Also, due to the range of changes and the scale of measurement of each of the studied parameters, the normalization of this data has a significant role in the data-driven system modeling approaches with appropriate accuracy, because if they are not normalized, a factor with a larger scale may cause a computational deviation. Therefore, all data is normalized using the min-max normalization before modelings.

		Actual	
		P	N
Predicted	P	TP	FP
	N	FN	TN

- *P is the number of real positive cases in the data.*
- *N is the number of real negative cases in the data.*
- *TP is the number of correct predictions that an instance is positive.*
- *TN is the number of correct predictions that an instance is negative.*
- *FP is the number of incorrect predictions that an instance is positive.*
- *FN is the number of incorrect of predictions that an instance negative.*

Figure 3. A basic form of a confusion matrix.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (24)$$

$$Error = \frac{FP + FN}{TP + FP + TN + FN} = 1 - Accuracy \quad (25)$$

4.1. GMDH modeling

In this work, GMDH was employed to construct a binary classification model for the assessment of the number of vehicles involved in the accident using the MATLAB environment. The number of vehicles involved in the accident was considered as the dependent variable, and DL, W, TA, L, SL, AS, and AADT were set as the independent variables. Determination of the optimal architecture of GMDH contributes greatly to the high performance of GMDH models, hence, accurate determination of control parameters of the GMDH model is an essential issue. Although there are no specific relationships to accurately determine control parameters, most of these parameters are obtained from past studies, expert opinion, and trial and error. Therefore, according to experts' opinions and previous studies, a range is considered for some control parameters of the GMDH model, which includes, the set of Maximum Number of Layers (MNL) equal to 5, 10, 20, 40, and 50, and Maximum Number of Neurons in a Layer (MNNL) equal to 5, 10, 20, 40 and 50. Also, the Selection Pressure (SP) is another control parameter of the GMDH model and it is a dimensionless number where the sensitivity of modeling error is affected by the SP and it was considered equal to 0.5. Different rates for training and testing data are considered in modeling and several studies have been conducted in this case [100-101]. In this modeling, based on the suggestions and studies of Looney, 75% of the data (423 cases) were used for training and 25% of the data (141 cases) were applied as test data for modeling [102]. In total, 25 models were constructed and their results are demonstrated in Table 3.

Table 3. The accuracies of training and testing of models based on the various control parameters.

Models No.	SP	MNL	MNNL	Training Accuracy (%)	Testing Accuracy (%)
1	0.5	5	5	78.5	78
2	0.5	5	10	79.9	73.8
3	0.5	5	20	79.7	74.5
4	0.5	5	40	80.1	73.8
5	0.5	5	50	80.9	77.3
6	0.5	10	5	78.3	75.2
7	0.5	10	10	77.8	76.6
8	0.5	10	20	77.8	76.6
9	0.5	10	40	79.2	75.9
10	0.5	10	50	82.7	80.1
11	0.5	20	5	78	75.9
12	0.5	20	10	80.9	73.8
13	0.5	20	20	79.9	77.3
14	0.5	20	40	82	80.9
15	0.5	20	50	83.2	81.6
16	0.5	40	5	79.9	77.3
17	0.5	40	10	80.1	78.7
18	0.5	40	20	82.7	78.7
19	0.5	40	40	81.3	79.4
20	0.5	40	50	80.1	76.6

21	0.5	50	5	77.8	76.6
22	0.5	50	10	80.9	74.5
23	0.5	50	20	77.8	76.6
24	0.5	50	40	78.3	75.2
25	0.5	50	50	82	77.3

After constructing different models and determining the accuracy performance of each model, all models were ranked based on a simple ranking method suggested by Zorlu et al., (2008) and Table 4 indicates the ranking results [103].

Table 4. Ranking of developed models.

Models No.	SP	MNL	MNNL	Rating for Training Accuracy	Rating for Testing Accuracy	Total Rank
1	0.5	5	5	16	20	36
2	0.5	5	10	19	14	33
3	0.5	5	20	18	15	33
4	0.5	5	40	20	14	34
5	0.5	5	50	21	19	40
6	0.5	10	5	15	16	31
7	0.5	10	10	13	18	31
8	0.5	10	20	13	18	31
9	0.5	10	40	17	17	34
10	0.5	10	50	24	23	47
11	0.5	20	5	14	17	31
12	0.5	20	10	21	14	35
13	0.5	20	20	19	19	38
14	0.5	20	40	23	24	47
15	0.5	20	50	25	25	50
16	0.5	35	5	19	19	38
17	0.5	35	10	20	21	41
18	0.5	35	20	24	21	45
19	0.5	35	40	22	22	44
20	0.5	35	50	20	18	38
21	0.5	50	5	13	18	31
22	0.5	50	10	21	15	36
23	0.5	50	20	13	18	31
24	0.5	50	40	15	16	31
25	0.5	50	50	23	19	42

According to Table 4, all models were ranked and the seven developed models had the lowest rank with a rank equal to 31 among 25 developed models. While the 15th model had the best performance in comparison with other developed models. Its training accuracy and testing accuracy were 83.2% and 81.6%. The structure of the 15th developed model consisted of MNL, MNNL, and SP equal to 20, 50 and 0.5, respectively. The results of confusion matrices of the training, the testing and total data are indicated in Fig. 4 (a), (b) and (c), respectively.

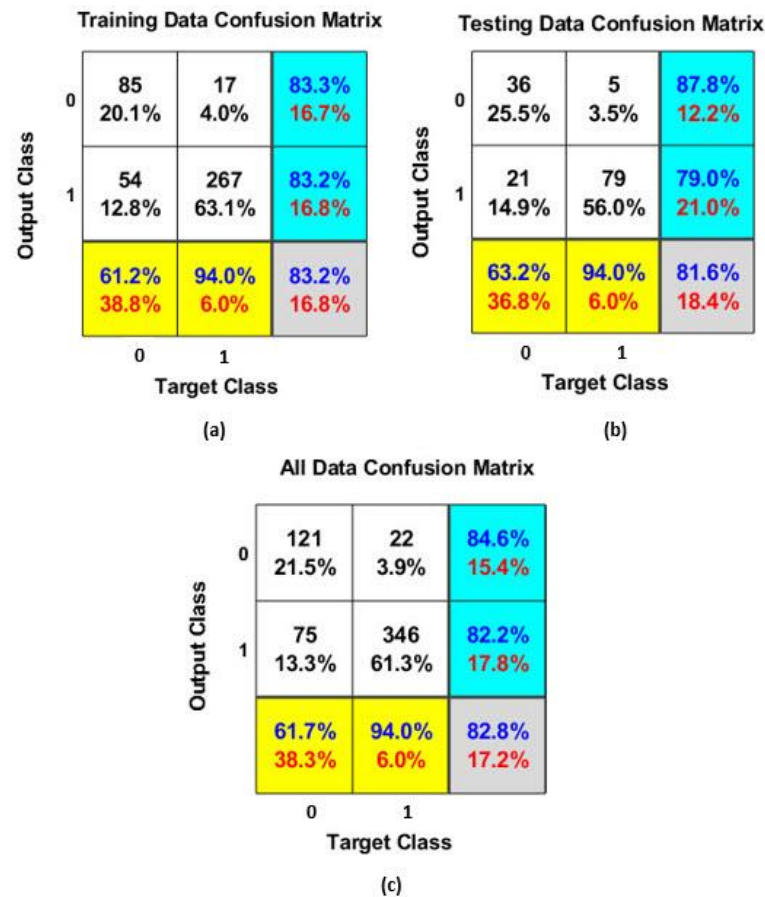


Figure 4. Confusion matrices of training data (a), testing data (b), and total data (c).

As mentioned earlier, 75% data set (423) were considered as the training dataset and the rest (141) were assigned for the testing dataset. According to Fig. 4 (a), the 15th binary classification developed model could correctly recognize 80 cases of the first class with label "0" (At most one car was involved in the accident), and 18 cases of the second class (At least two or more vehicles were involved in an accident) were wrongly estimated in the first class with label "0". Note that, it could classify the training dataset with an accuracy equal to 83.2%. Meanwhile, the results of binary classification for the testing data according to Fig. 4 (b), show that 36 cases of the first class with label "0" and 79 cases of the second class with label "1" were correctly considered, while the 15th developed model wrongly predicted 5 and 21 cases of the first and second classes with labels "0" and "1", respectively. Consequently, the 15th model was able to obtain an acceptable accuracy of classifying test data with 81.6%. Finally, according to the binary classification results for the whole data according to Fig. 4 (c), it is clear that the 15th binary classification developed model could correctly classified 467 cases of two classes, and 97 cases of both classes were wrongly classified. Consequently, the total accuracy of binary classification modeling was 82.8%. This analysis shows that GMDH is a reliable modeling method for predicting the number of accidents.

4.2. GOA-SVM modeling

A combination of GOA and SVM was used to develop a predictive model in the MATLAB environment. The GOA algorithm was used to optimize some parameters of the SVM so that the SVM model shows the highest performance. In the GOA-SVM modeling of this study, the same datasets performed in the analysis of GMDH were used. For modelling, 75% of the dataset (423 data) is randomly defined as a training dataset and the remaining 25% (141 data) is considered as the testing dataset [102]. Like modeling with the GMDH algorithm, the two classes were considered for all data with labels including

"0" and "1". To develop and optimize the parameters of the SVM model by the GOA, the control parameters of the GOA must be determined, which play an important role in the rapid and appropriate convergence of the model. Although there are no specific relationships to determine these parameters, based on previous studies and experts' opinions, a range was determined for each of them, such as the number of grasshoppers' populations (5,10,15,20,30 and 40) and the number of iterations (10, 20, 40, 50 and 100). Then, the most appropriate ones were selected by trial and error approach [104].

Also, to further evaluate the model, the k-fold cross-validation was used, in which the data was subdivided into K subsets. In this system, one of them at each time for validation, while other K-1 for training were applied. This procedure was carried out K times, with each data set being used exactly once for training and once for validation. Finally, the average result of this K validation was chosen as a final estimate. There is no specific method for determining the amount of k-fold and it is determined according to the number of data and the opinion of experts, hence, k-fold was considered equal to 3 in this study [104]. In addition, the three different types of kernel functions including RBF, POL, and LIN were used. According to the number of control parameters, 30 models were built for each kernel function of GOA-SVM were made that in total, 90 models were made for GOA-SVM. Preliminary analyzes were performed and a comparison of the results of the performance indicators for the best-developed models with three different kernels is shown in Fig. 5.

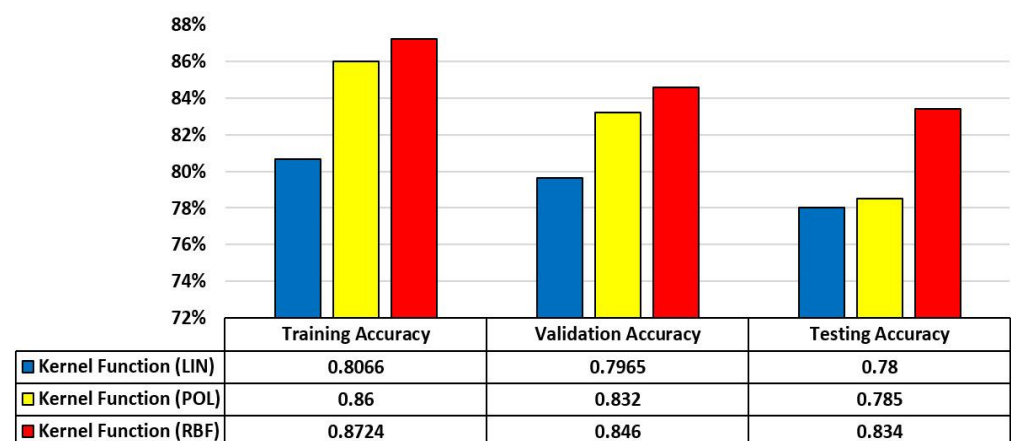


Figure 5. Comparison between the values of training accuracy, validation accuracy, and testing accuracy for the best developed models.

Although, the GOA algorithm was able to train SVM very well with different kernels according to Fig. 5, it is clear that the best-developed model with RBF kernel had the highest amount of accuracy in training, validation, and testing in comparison with POL and LIN kernel functions. Therefore, the best developed GOA-SVM model is considered with RBF kernel function and the optimal control parameters of the model are shown in Table 5. Also, the value of error in each iteration was calculated based on Eq 25 and the result is indicated in Fig. 6. According to Fig. 6, modeling started with an error of about 0.195 and had different values until the 18th iteration and reached 0.153 in the 19th iteration, which remained constant until the last iteration (40th). The obtained values of accuracy and the error indicated the proper convergence of the model.

Table 5. Control parameters of the best developed GOA-SVM model.

No	Control parameters	Values
1	Grasshoppers' populations	40
2	Number of iterations	40
3	k-fold	3

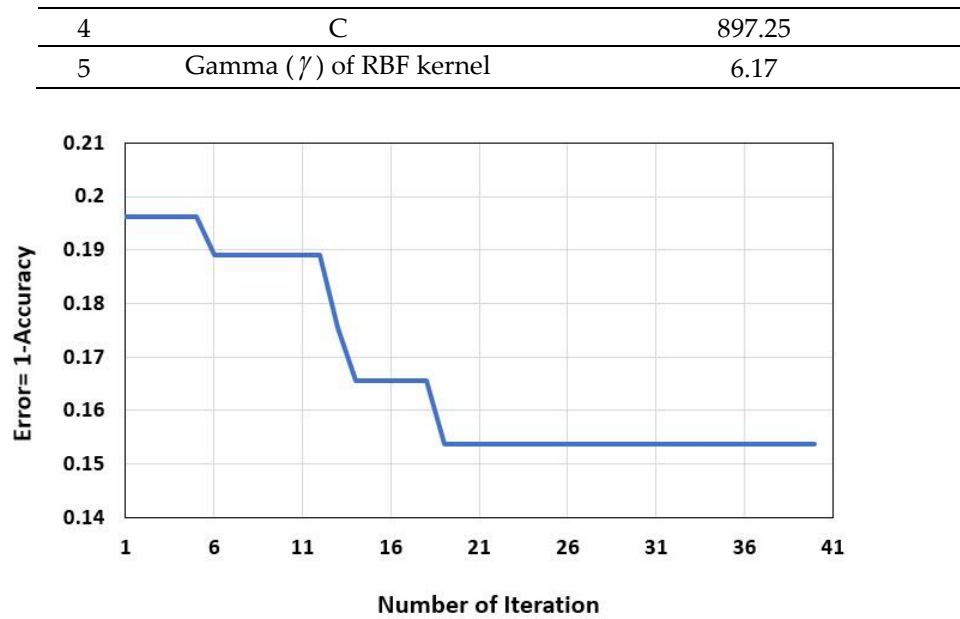


Figure 6. Variations of error value in each iteration during GOA-SVM modeling.

4.3. Comparison of models' performance and sensitivity analysis

Each year, many people pass away as a result of road traffic accidents. therefore, knowing the impact of various contributing factors on road accidents and taking the necessary measures to reduce accidents can have a significant impact on increasing the level of road safety. In this research, two machine learning methods, namely GMDH-type neural network and GOA-SVM were employed to conduct the binary classification modeling. After multiple modeling, the best GMDH model and GOA-SVM model were selected based on the accuracy of the modeling performance. A comparison was made between the best model of GMDH and the best model of GOA-SVM based on the accuracy of training and testing that is shown in Fig. 7. According to explanations in the previous section, it should be noted that the value of the validation accuracy model is considered instead of the value of training accuracy in the GOA-SVM model, hence the value of validation accuracy of the GOA-SVM model should be compared to the value of training accuracy of the GMDH model.

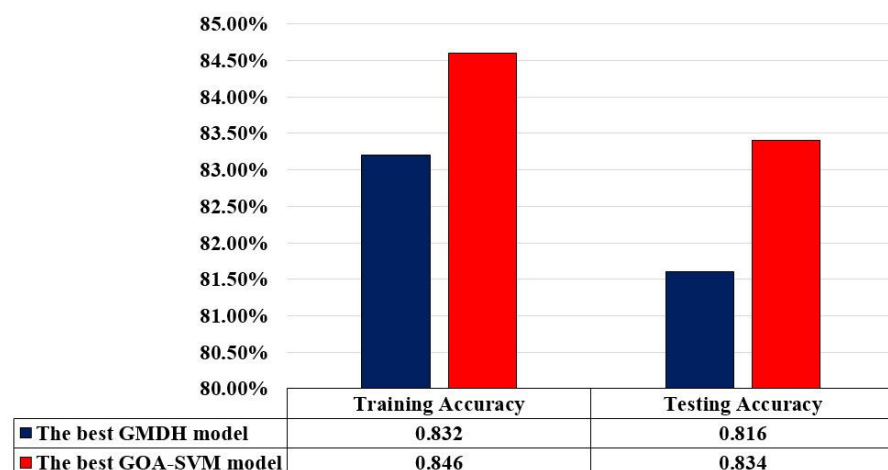


Figure 7. Comparison between the best GMDH and GOA-SVM models for training and testing accuracies.

According to Fig. 7., the best GOA-SVM model indicated higher performance than the best GMDH model for predicting the number of accidents by 84.6% and 83.4% for training and testing accuracies in comparison with 83.2% and 81.6% for training and testing accuracies, respectively. Although, it is worth mentioning that both models had acceptable degrees of accuracy and robustness, therefore, it can be concluded that they are reliable systems of modeling for predicting the number of crashes and can be used as useful tools for modeling roads safety involved in transportation engineering.

Road accidents can cause considerable economic and human losses to society, therefore, assessing the impact of parameters affecting the number of accidents can provide an in-depth insight for engineers involved in road safety management. Consequently, to assess the impact of DL, W, TA, L, SL, AS, and AADT on the number of predicted vehicles involved in the accident, a sensitivity analysis was performed. This sensitivity analysis is based on the cosine amplitude method according to Eq 26 which r_{ij} is the strength of the relation, n shows the number of dataset and also x_{ik} and y_{ij} explain input variables and the predicted output, correspondingly [104-105].

$$r_{ij} = \frac{\sum_{k=1}^n (x_{ik} \times y_{jk})}{\sqrt{\sum_{k=1}^n x_{ik}^2 \sum_{k=1}^n y_{jk}^2}} \quad (26)$$

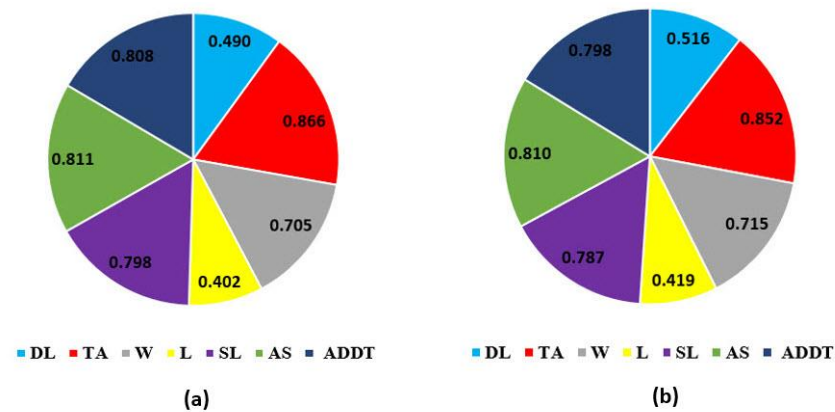


Figure 8. Comparison between the results of sensitivity analysis obtained from GMDH model (a) and GOA-SVM model (b).

According to Fig. 8, it is clear that the analyses of both models had similar results from the impact of the factors under consideration, which indicates the reliability of the results. Furthermore, the following remarks can be concluded:

- Type of accident was the most significant factor among other contributing factors that affect the number of vehicles involved in the crashes. Generally, there are some types of accidents such as head-on, road departure, rear-end, side collisions that several problems such as the lack of traffic signs, road quality, and driver conditions can affect them. This factor has a important effect on the severity of injuries and fatalities. Therefore, road safety can increase in rural roads of Cosenza in southern Italy by providing geometric conditions such as improvement of roadway or guideway geometry, topography, and roadway pavement surfaces.

- The next factor is the average speed, which can increase the risk of accidents. An increase in the average speed leads to an increase in the severity of the injury, although, some researchers found out that the effects of average speed on the crash rate can be reduced and ignored by controlling other factors affecting crashes for instance traffic volume, road geometry, and the number of lanes [106-107]. When the average speed is higher, the driver's response time is shorter, which this issue can lead to an accident especially on

holidays. Therefore, it is possible to control the impact of average speed by providing some types of measures such as improvement of the location of road signs, speed limit enforcement methods, pavement markings, and vertical centerline treatments.

- The third factor influencing the number of vehicles involved in the accident after the type of accident is the Annual average daily traffic that plays a key role in the development needs and priorities of roads development for transportation planning. Also, some studies indicate that increasing the amount of AADT can lead to an increase in the frequency of accidents [108-109]. Therefore, to reduce the effects of AADT on this case study, it is recommended for planning new roads or considering other intercity transportation systems such as trains which are considered in the review of coming urban development plans.

- The subsequent contributing factor is speed limit that has a significant role in the behavior and decision of drivers. Generally, it is considered that the speed limit is determined by the road conditions. If the speed limit is selected incorrectly in a part of the route and the driver is aware of this error due to the road conditions, they may lose confidence in the speed limits in other sections of the road and increase or decrease the speed based on their interpretation [110-111]. Hence, given the impact of this factor in this case study, it is suggested that a general review be considered in selecting the speed limit for rural roads of Cosenza.

- Several appropriate studies have been done on the relationship between weekdays and accidents [112-113]. In this study, out of 564 accidents, 367 accidents occurred on holidays, which about 65% of accidents occurred on holidays. Due to the geographical location of Cosenza, the amount of traffic on holidays has a relative increase, therefore, controlling and monitoring more rural roads such as using automated speed enforcement and considering stricter laws with adequate penalties can expect to reduce the effect of this factor.

- Extensive studies have been conducted on the effect of daylight on the number of vehicles involved in the accident, which shows the high importance of this factor. In many studies, this factor has been given the first importance among other factors [48], [114]. Geographical location and road lighting systems play a significant role in the amount of impact of this factor. This factor was determined as the sixth effective factor out of seven factors, which this result is matched with the location and road lighting system of rural roads of Cosenza.

- The last studied factor was the location that had the least impact on the rate of vehicles involved in the accident based on the results of both artificial intelligence models. It can be concluded that based on the type of structure of rural roads of Cosenza, the Location has not had much effect on the rate of accidents. Therefore, in future studies on this case study, the effect of this factor can be ignored.

Therefore, it can be concluded that several measures such as improving roadway geometry, pavement markings, and vertical centerline treatments are the most significant measures that should be considered in reviewing the plan for rural roads of Cosenza. Finally, it should be noted that the presented models with specific structures are location-sensitive and cannot directly be used in other rural roads.

5. Conclusions

Evaluation and analysis of contributing factors affecting the number of crashes can lead to a deep understanding of the existing situation and increase the level of road safety by planning and taking a series of necessary measures. Therefore, an attempt was made to predict the number of vehicles involved in the crashes using the GMDH-type neural network and GOA-SVM methods in this study. This study was accomplished using 546 accident cases of rural roads of Cosenza in southern Italy. The several influential parameters on crashes, namely DL, W, TA, L, SL, AS, and AADT were set as inputs variables and the number of vehicles involved in the crash was considered as output data. According to the number of control parameters of each technique, the 25 models were made for

GMDH and the 30 models for each kernel function of GOA-SVM that the 90 models were made for GOA-SVM in total. After modeling, the made models in each technique were compared with each other in terms of the model's performance. Among the binary classification of GMDH models, the 15th model was selected with the highest score and the GOA-SVM model with RBF kernel function was chosen among the binary classification of GOA-SVM models. In comparison between the GMDH and GOA-SVM models, the GOA-SVM model had the higher capability in the prediction of the number of vehicles involved in the accident, although the performance of both models indicates that both can be used as useful tools for modeling the number of vehicles involved in the accident in transportation engineering. Consequently, a sensitivity analysis was performed based on the results obtained from both models. In both models, the Type of accident and Location had the highest and lowest impact compared to other parameters on the rate of accidents on rural roads of Cosenza, respectively. For future work, it is recommended to examine other factors that can have an impact on the number of vehicles involved in the accidents such as the age of drivers, the age of cars, gender, and the geometry of roads.

Author Contributions: Conceptualization, G.G. and S.S.H. (Sina Shaffiee Haghshenas); methodology, S.S.H. (Sina Shaffiee Haghshenas) and S.S.H. (Sami Shaffiee Haghshenas); software, S.S.H. (Sina Shaffiee Haghshenas); formal analysis, S.S.H. (Sina Shaffiee Haghshenas); investigation, S.S.H. (Sina Shaffiee Haghshenas) and S.S.H. (Sami Shaffiee Haghshenas); resources, G.G., V.A., A.V.; writing—original draft preparation, S.S.H. (Sina Shaffiee Haghshenas); writing—review and editing, G.G., S.S.H. (Sina Shaffiee Haghshenas), S.S.H. (Sami Shaffiee Haghshenas), V.A., Y.P., Z.W.G.; supervision, G.G., V.A., A.V., Y.P., Z.W.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We would like to express our deepest thanks to Mahdi Ghaem for his excellent advice

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Nordfjærn, T., & Rundmo, T. Road traffic safety beliefs and driver behaviors among personality subtypes of drivers in the Norwegian population. *Traffic Inj. Prev.*, **2013**, 14(7), 690-696.
2. Ellison, A. B., Greaves, S. P., & Bliemer, M. C. Driver behaviour profiles for road safety analysis. *Accid Anal Prev*, **2015**, 76, 118-132.
3. Scott-Parker, B. Emotions, behaviour, and the adolescent driver: A literature review. *Transp. Res. F: Traffic Psychol. Behav*, **2017**, 50, 1-37.
4. Farooq, D., Moslem, S., & Duleba, S. Evaluation of driver behavior criteria for evolution of sustainable traffic safety. *Sustainability*, **2019**, 11(11), 3142.
5. Farooq, D., & Juhasz, J. Statistical Evaluation of Risky Driver Behavior Factors That Influence Road Safety Based on Drivers Age and Driving Experience in Budapest and Islamabad. *Eur. Transp. Eur*, **2020**, 1-18.
6. Rosli, N., Ambak, K., Shahidan, N. N., Sukor, N. S. A., Osman, S., & Yei, L. S. Driving behaviour of elderly drivers in Malaysia. *Int. J. Integr. Eng.*, **2020**, 12(8), 268-277.
7. Christoforou, Z. D., Karlaftis, M. G., & Yannis, G. (2010, March). Heavy vehicle age and road safety. In *Proceedings of the Institution of Civil Engineers-Transport* (Vol. 163, No. 1, pp. 41-48). Thomas Telford Ltd.
8. Russo, F., Biancardo, S. A., & Dell'Acqua, G. Road safety from the perspective of driver gender and age as related to the injury crash frequency and road scenario. *Traffic Inj. Prev.*, **2014**, 15(1), 25-33.
9. Kim, S., & Kim, J. K. Road safety for an aged society: compliance with traffic regulations, knowledge about traffic regulations, and risk factors of older drivers. *Transp. Res. Rec*, **2017**, 2660(1), 15-21.

10. Casado-Sanz, N., & Guirao, B. Analysis of the impact of population ageing and territorial factors on crosstown roads safety: the Spanish case study. *Transp. Res. Procedia*, **2018**, 33, 283-290.
11. Török, Á. A Novel Approach in Evaluating the Impact of Vehicle Age on Road Safety. *PROMET-ZAGREB*, **2020**, 32(6), 789-796.
12. Lyon, C., Mayhew, D., Granie, M. A., Robertson, R., Vanlaar, W., Woods-Fry, H., ... & Soteropoulos, A. Age and road safety performance: focusing on elderly and young drivers. *IATSS research*, **2020**, 44(3), 212-219.
13. Theofilatos, A., & Yannis, G. A review of the effect of traffic and weather characteristics on road safety. *Accid Anal Prev*, **2014**, 72, 244-256.
14. Lazarev, Y., Medres, C., Raty, J., & Bondarenko, A. Method of assessment and prediction of temperature conditions of roadway surfacing as a factor of the road safety. *Transp. Res. Procedia*, **2017**, 20, 393-400.
15. Malin, F., Norros, I., & Innamaa, S. Accident risk of road and weather conditions on different road types. *Accid Anal Prev*, **2019**, 122, 181-188.
16. Ivajnsič, D., Horvat, N., Žiberna, I., Kotnik, E. K., & Davidović, D. Revealing the Spatial Pattern of Weather-Related Road Traffic Crashes in Slovenia. *Appl Sci*, **2021**, 11(14), 6506.
17. Abdella, G. M., & Shaaban, K. Modeling the impact of weather conditions on pedestrian injury counts using LASSO-based poisson model. *Arab J Sci Eng*, **2021**, 46(5), 4719-4730.
18. Bajwa, S., Warita, M., & Kuwahara, M. (2010). Effects of road geometry, weather and traffic flow on road safety. In Proceedings of the 15th International Conference of Hong Kong Society for Transportation Study, 773-780.
19. Orfila, O., Coiret, A., Do, M. T., & Mammar, S. Modeling of dynamic vehicle-road interactions for safety-related road evaluation. *Accid Anal Prev*, **2010**, 42(6), 1736-1743.
20. Alian, S., Baker, R. G. V., & Wood, S. Rural casualty crashes on the Kings Highway: A new approach for road safety studies. *Accid Anal Prev*, **2016**, 95, 8-19.
21. Ewan, L., Al-Kaisy, A., & Hossain, F. Safety effects of road geometry and roadside features on low-volume roads in Oregon. *Transp. Res. Rec*, **2016**, 2580(1), 47-55.
22. Gooch, J. P., Gayah, V. V., & Donnell, E. T. Quantifying the safety effects of horizontal curves on two-way, two-lane rural roads. *Accid Anal Prev*, **2016**, 92, 71-81.
23. Roslak, J., & Wallaschek, J. (2004, June). Active lighting systems for improved road safety. In IEEE Intelligent Vehicles Symposium, 682-685. (2004, June).
24. Magar, S. G. Adaptive Front Light Systems of Vehicle for Road Safety. In 2015 International Conference on Computing Communication Control and Automation, 551-554. (2015, February).
25. Aldulaimi, M., & AmadorJimenez, L. (2016). 759 Road lighting and safety: a pilot study of Arthabaska region. 22: A272.
26. Tetervenoks, O., Avotins, A., Fedorjana, N., Kluga, A., & Krasts, V. Potential Role of Street Lighting System for Safety Enhancement on the Roads in Future. In 2019 IEEE 60th International Scientific Conference on Power and Electrical Engineering of Riga Technical University (RTUCON), 1-5. (2019, October).
27. Liu, J., Li, J., Wang, K., Zhao, J., Cong, H., & He, P. Exploring factors affecting the severity of night-time vehicle accidents under low illumination conditions. *Adv. Mech. Eng*, **2019**, 11(4), 1687814019840940.
28. Saljoqi, M., Behnood, H. R., & Mirbaha, B. Developing the crash modification model for urban street lighting. *Innov. Infrastruct. Solut*, **2021**, 6(2), 1-10.
29. Zou, X., Yue, W. L., & Le Vu, H. Visualization and analysis of mapping knowledge domain of road safety studies. *Accid Anal Prev*, **2018**, 118, 131-145.
30. Elvik, R., Vadeby, A., Hels, T., & van Schagen, I. Updated estimates of the relationship between speed and road safety at the aggregate and individual levels. *Accid Anal Prev*, **2019**, 123, 114-122.
31. Gichaga, F. J. The impact of road improvements on road safety and related characteristics. *IATSS research*, **2017**, 40(2), 72-75.
32. Sager, L. Estimating the effect of air pollution on road safety using atmospheric temperature inversions. *J Environ Econ Manage*, **2019**, 98, 102250.
33. Mahmud, S. M. S., Ferreira, L., & Tavassoli, A. (2016, May), Traditional approaches to Traffic Safety Evaluation (TSE): application challenges and future directions. 11th Asia Pacific Transportation Development Conference and 29th ICTPA Annual Conference, 242-262.
34. Zheng, L., Ismail, K., Sayed, T., & Fatema, T. Bivariate extreme value modeling for road safety estimation. *Accid Anal Prev*, **2018**, 120, 83-91.
35. Wang, C., Xu, C., & Dai, Y. A crash prediction method based on bivariate extreme value theory and video-based vehicle trajectory data. *Accid Anal Prev*, **2019**, 123, 365-373.
36. Zheng, L., Sayed, T., & Mannering, F. Modeling traffic conflicts for use in road safety analysis: a review of analytic methods and future directions. *Anal. Methods Accid*, **2020**, 100142.
37. Jamal, A., Rahman, M. T., Al-Ahmadi, H. M., & Mansoor, U. The dilemma of road safety in the eastern province of Saudi Arabia: Consequences and prevention strategies. *Int. J. Environ. Res. Public Health*, **2020**, 17(1), 157.
38. Mussone, L., Ferrari, A., & Oneta, M. An analysis of urban collisions using an artificial intelligence model. *Accid Anal Prev*, **1999**, 31(6), 705-718.
39. Halim, Z., Kalsoom, R., Bashir, S., & Abbas, G. Artificial intelligence techniques for driving safety and vehicle crash prediction. *Artif. Intell. Rev*, **2016**, 46(3), 351-387.
40. Castro, Y., & Kim, Y. J. Data mining on road safety: factor assessment on vehicle accidents using classification models. *Int. J. Crashworthiness*, **2016**, 21(2), 104-111.

41. De Luca, M. A comparison between prediction power of artificial neural networks and multivariate analysis in road safety management. *Transport*, **2017**, 32(4), 379-385.
42. Shah, S. A. R., Brijis, T., Ahmad, N., Pirdavani, A., Shen, Y., & Basheer, M. A. Road safety risk evaluation using gis-based data envelopment analysis—artificial neural networks approach. *Appl Sci*, **2017**, 7(9), 886.
43. Liu, M., Wu, Z., Chen, Y., & Zhang, X. (2020). Utilizing Decision Tree Method and ANFIS to Explore Real-Time Crash Risk for Urban Freeways. In CICTP 2020, 2495-2508.
44. Guido, G., Haghshenas, S. S., Haghshenas, S. S., Vitale, A., Gallelli, V., & Astarita, V. Development of a binary classification model to assess safety in transportation systems using GMDH-type neural network algorithm. *Sustainability*, **2020**, 12(17), 6735.
45. Mokhtarimousavi, S., Anderson, J. C., Hadi, M., & Azizinamini, A. A temporal investigation of crash severity factors in worker-involved work zone crashes: Random parameters and machine learning approaches. *Transp. Res. Interdiscip. Perspect*, **2021**, 10, 100378.
46. Kitali, A. E., Mokhtarimousavi, S., Kadeha, C., & Alluri, P. Severity analysis of crashes on express lane facilities using support vector machine model trained by firefly algorithm. *Traffic Inj. Prev.*, **2021**, 22(1), 79-84.
47. Xu, Y., Ye, Z., Wang, Y., Wang, C., & Sun, C. Evaluating the influence of road lighting on traffic safety at accesses using an artificial neural network. *Traffic Inj. Prev.*, **2018**, 19(6), 601-606.
48. Amiri, A. M., Sadri, A., Nadimi, N., & Shams, M. A comparison between artificial neural network and hybrid intelligent genetic algorithm in predicting the severity of fixed object crashes among elderly drivers. *Accid Anal Prev*, **2020**, 138, 105468.
49. Guido, G., Haghshenas, S. S., Haghshenas, S. S., Vitale, A., Astarita, V., & Haghshenas, A. S. Feasibility of stochastic models for evaluation of potential factors for safety: a case study in Southern Italy. *Sustainability*, **2020**, 12(18), 7541.
50. Shiran, G., Imaninasab, R., & Khayamim, R. Crash Severity Analysis of Highways Based on Multinomial Logistic Regression Model, Decision Tree Techniques, and Artificial Neural Network: A Modeling Comparison. *Sustainability*, **2021**, 13(10), 5670.
51. Mussone, L., Bassani, M., Masci, P. Analysis of factors affecting the severity of crashes in urban road intersections. *Accid Anal Prev*, **2017**, 103, pp. 112-122, 2017.
52. Ministero delle infrastrutture e dei Trasporti, Nuovo Codice della Strada, Decreto Legislativo N. 285 del 30/4/1992, G.U. n. 114 del 18/5/1992.
53. Ministero delle infrastrutture e dei Trasporti, Disposizioni urgenti per la sicurezza della circolazione dei veicoli e di specifiche categorie di utenti, Modifiche al Nuovo Codice della Strada, Decreto Legislativo n.121 del 10/9/2021, G.U. n. 267 del 9/9/2021.
54. Pan, G., Fu, L., & Thakali, L. Development of a global road safety performance function using deep neural networks. *Int. J. Transp. Sci. Technol*, **2017**, 6(3), 159-173.
55. Peng, Z., Gao, S., Li, Z., Xiao, B., & Qian, Y. Vehicle safety improvement through deep learning and mobile sensing. *IEEE Netw*, **2018**, 32(4), 28-33.
56. Silva, P. B., Andrade, M., & Ferreira, S. Machine learning applied to road safety modeling: a systematic literature review. *Journal of traffic and transportation engineering (English edition)*, **2020**, 7(6), 775-790.
57. Hernández, H., Alberdi, E., Pérez-Acebo, H., Álvarez, I., García, M. J., Eguia, I., & Fernández, K. Managing Traffic Data through Clustering and Radial Basis Functions. *Sustainability*, **2021**, 13(5), 2846.
58. Amiri, A. M., Naderi, K., Cooper, J. F., & Nadimi, N. (2021). Evaluating the impact of socio-economic contributing factors of cities in California on their traffic safety condition. *J. Transp. Health*, **2021**, 20, 101010.
59. Jamal, A., Zahid, M., Tauhidur Rahman, M., Al-Ahmadi, H. M., Almoshaogeh, M., Farooq, D., & Ahmad, M. Injury severity prediction of traffic crashes with ensemble machine learning techniques: a comparative study. *Int J Inj Contr Saf Promot*, **2021**, 1-20.
60. Tehrany, M. S., Pradhan, B., & Jebur, M. N. (2014). Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. *J. Hydrol*, **2014**, 512, 332-343.
61. Jahed Armaghani, D., Asteris, P. G., Askarian, B., Hasanipanah, M., Tarinejad, R., & Huynh, V. V. Examining hybrid and single SVM models with different kernels to predict rock brittleness. *Sustainability*, **2020**, 12(6), 2229.
62. Mikaeil, R., Haghshenas, S. S., Ozcelik, Y., & Gharehgheshlagh, H. H. Performance evaluation of adaptive neuro-fuzzy inference system and group method of data handling-type neural network for estimating wear rate of diamond wire saw. *Geotech. Geol. Eng.*, **2018**, 36(6), 3779-3791.
63. Dormishi, A. R., Ataei, M., Khaloo Kakaie, R., Mikaeil, R., & Shaffiee Haghshenas, S. Performance evaluation of gang saw using hybrid ANFIS-DE and hybrid ANFIS-PSO algorithms. *JME*, **2019**, 10(2), 543-557.
64. Naderpour, H., & Mirrashid, M. Moment capacity estimation of spirally reinforced concrete columns using ANFIS. *Complex intell. systems*, **2020**, 6(1), 97-107.
65. Faradonbeh, R. S., Haghshenas, S. S., Taheri, A., & Mikaeil, R. Application of self-organizing map and fuzzy c-mean techniques for rockburst clustering in deep underground projects. *Neural. Comput. Appl*, **2020**, 32(12), 8545-8559.
66. Noori, A. M., Mikaeil, R., Mokhtarian, M., Haghshenas, S. S., & Foroughi, M. Feasibility of intelligent models for prediction of utilization factor of TBM. *Geotech. Geol. Eng.*, **2020**, 38(3), 3125-3143.
67. Golafshani, E. M., & Behnood, A. Predicting the mechanical properties of sustainable concrete containing waste foundry sand using multi-objective ANN approach. *Constr Build Mater*, **2021**, 291, 123314.
68. Ivakhnenko, A.G. Polynomial Theory of Complex Systems. *IEEE Trans. Syst. Man Cybern.* 1971, SMC-1, 364-378, doi:10.1109/tsmc.1971.4308320.
69. Sezavar, R., Shafabakhsh, G., & Mirabdolazimi, S. M. New model of moisture susceptibility of nano silica-modified asphalt concrete using GMDH algorithm. *Constr Build Mater*, **2019**, 211, 528-538.

70. Dag, O., Kasikci, M., Karabulut, E., & Alpar, R. Diverse classifiers ensemble based on GMDH-type neural network algorithm for binary classification. *Commun. Stat. Simul. Comput*, **2019**, 1-17.
71. Armaghani, D. J., Hasanipanah, M., Amnieh, H. B., Bui, D. T., Mehrabi, P., & Khorami, M. Development of a novel hybrid intelligent model for solving engineering problems using GS-GMDH algorithm. *Eng Comput*, **2019**, 36,1379–1391.
72. Harandizadeh, H., Armaghani, D. J., & Mohamad, E. T. Development of fuzzy-GMDH model optimized by GSA to predict rock tensile strength based on experimental datasets. *Neural. Comput. Appl*, **2020**, 32(17), 14047-14067.
73. Jeddi, S., & Sharifian, S. A hybrid wavelet decomposer and GMDH-ELM ensemble model for Network function virtualization workload forecasting in cloud computing. *Appl. Soft Comput*, **2020**, 88, 105940.
74. Fiorini Morosini, A., Shaffiee Haghsheenas, S., Shaffiee Haghsheenas, S., & Geem, Z. W. Development of a binary model for evaluating water distribution systems by a pressure driven analysis (PDA) approach. *Appl Sci*, **2020**, 10(9), 3029.
75. Khandezamin, Z., Naderan, M., & Rashti, M. J. Detection and classification of breast cancer using logistic regression feature selection and GMDH classifier. *J. Biomed. Inform*, **2020**, 111, 103591.
76. Fiorini Morosini, A., Shaffiee Haghsheenas, S., Shaffiee Haghsheenas, S., Choi, D. Y., & Geem, Z. W. Sensitivity Analysis for Performance Evaluation of a Real Water Distribution System by a Pressure Driven Analysis Approach and Artificial Intelligence Method. *Water*, **2021**, 13(8), 1116.
77. Pusat, S., & Akkaya, A. V. Explicit equation derivation for predicting coal moisture content in convective drying process by GMDH-type neural network. *Int. J. Coal Prep*, **2020**, 1-14.
78. Vissol-Gaudin, E., Kotsialos, A., Groves, C., Pearson, C., Zeze, D. A., & Petty, M. C. (2017, November). Computing based on material training: Application to binary classification problems. In 2017 IEEE International Conference on Rebooting Computing (ICRC), 1-8. IEEE.
79. Li, D., Armaghani, D. J., Zhou, J., Lai, S. H., & Hasanipanah, M. A GMDH predictive model to predict rock material strength using three non-destructive tests. *J. Nondestruct. Eval*, **2020**, 39(4), 1-14.
80. Cortes, C., & Vapnik, V. Support-vector networks. *Mach Learn*, **1985**, 20(3), 273-297.
81. Chen, W. H., Hsu, S. H., & Shen, H. P. Application of SVM and ANN for intrusion detection. *Comput. Oper. Res*, **2005**, 32(10), 2617-2634.
82. Yan, X., & Jia, M. A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing. *Neural. Comput. Appl*, **2018**, 313, 47-64.
83. Zhou, S., Chu, X., Cao, S., Liu, X., & Zhou, Y. Prediction of the ground temperature with ANN, LS-SVM and fuzzy LS-SVM for GSHP application. *Geothermics*, **2020**, 84, 101757.
84. Maldonado, S., López, J., Jimenez-Molina, A., & Lira, H. Simultaneous feature selection and heterogeneity control for SVM classification: An application to mental workload assessment. *Expert Syst. Appl*, **2020**, 143, 112988.
85. Zeng, J., Roussis, P. C., Mohammed, A. S., Maraveas, C., Fatemi, S. A., Armaghani, D. J., & Asteris, P. G. Prediction of Peak Particle Velocity Caused by Blasting through the Combinations of Boosted-CHAID and SVM Models with Various Kernels. *Appl Sci*, **2021**, 11(8), 3705.
86. Zhou, J., Qiu, Y., Zhu, S., Armaghani, D. J., Li, C., Nguyen, H., & Yagiz, S. Optimization of support vector machine through the use of metaheuristic algorithms in forecasting TBM advance rate. *Eng. Appl. Artif. Intell*, **2021**, 97, 104015.
87. Mikaeil, R., Haghsheenas, S. S., & Hoseinie, S. H. Rock penetrability classification using artificial bee colony (ABC) algorithm and self-organizing map. *Geotech. Geol. Eng*, **2018**, 36(2), 1309-1318.
88. Salemi, A., Mikaeil, R., & Haghsheenas, S. S. (2018). Integration of finite difference method and genetic algorithm to seismic analysis of circular shallow tunnels (Case study: Tabriz urban railway tunnels). *KSCE J. Civ. Eng*, **2018**, 22(5), 1978-1990.
89. Aryafar, A., Mikaeil, R., Haghsheenas, S. S., & Haghsheenas, S. S. Application of metaheuristic algorithms to optimal clustering of sawing machine vibration. *MEAS*, **2018**, 124, 20-31.
90. Mikaeil, R., Haghsheenas, S. S., & Sedaghati, Z. Geotechnical risk evaluation of tunneling projects using optimization techniques (case study: the second part of Emamzade Hashem tunnel). *Nat. Hazards*, **2019**, 97(3), 1099-1113.
91. Mikaeil, R., Bakhshinezhad, H., Haghsheenas, S. S., & Ataei, M. Stability analysis of tunnel support systems using numerical and intelligent simulations (case study: Kouhin Tunnel of Qazvin-Rasht Railway). *Rud. Geol. Naft. Zb*, **2019**, 34(2), 1-10.
92. Haghsheenas, S. S., Faradonbeh, R. S., Mikaeil, R., Haghsheenas, S. S., Taheri, A., Saghatforoush, A., & Dormishi, A. A new conventional criterion for the performance evaluation of gang saw machines. *MEAS*, **2019**, 46, 159-170.
93. Saremi, S., Mirjalili, S., & Lewis, A. Grasshopper optimisation algorithm: theory and application. *Adv. Eng. Softw*, **2017**, 105, 30-47.
94. Mafarja, M., Aljarah, I., Faris, H., Hammouri, A. I., Ala'M, A. Z., & Mirjalili, S. Binary grasshopper optimisation algorithm approaches for feature selection problems. *Expert Syst. Appl*, **2019**, 117, 267-286.
95. Saxena, A. (2019). A comprehensive study of chaos embedded bridging mechanisms and crossover operators for grasshopper optimisation algorithm. *Expert Syst. Appl*, **2019**, 132, 166-188.
96. Pan, J. S., Wang, X., Chu, S. C., & Nguyen, T. T. A multi-group grasshopper optimisation algorithm for application in capacitated vehicle routing problem. *Pattern Recognit*, **2020**, 4(1), 41-56.
97. Goodarzizad, P., Mohammadi Golafshani, E., & Arashpour, M. Predicting the construction labour productivity using artificial neural network and grasshopper optimisation algorithm. *Int. J. Constr. Manag*, **2021**, 1-17.
98. Meraihi, Y., Gabis, A. B., Mirjalili, S., & Ramdane-Cherif, A. Grasshopper optimization algorithm: theory, variants, and applications. *IEEE Access*, **2021**, 9, 50001-50024.
99. Nelson, M. M., & Illingworth, W. T. (1991). A Practical Guide to Neural Nets, Addison-Wesley. Reading, MA.

100. Swingler, K. (1996) Applying Neural Networks: A Practical Guide. Academic, London, UK.
101. Looney, C. G. Advances in feedforward neural networks: demystifying knowledge acquiring black boxes. *IEEE Trans Knowl Data Eng*, **1996**, 8(2), 211-226.
102. Zorlu, K., Gokceoglu, C., Ocakoglu, F., Nefeslioglu, H. A., & Acikalin, S. J. E. G. Prediction of uniaxial compressive strength of sandstones using petrography-based models. *Eng. Geol*, **2008**, 96(3-4), 141-158.
103. Ataei, M., Mohammadi, S., & Mikaeil, R. Evaluating performance of cutting machines during sawing dimension stones. *J. Cent. South Univ*, **2019**, 26(7), 1934-1945.
104. Quddus, M. Exploring the relationship between average speed, speed variation, and accident rates using spatial statistical models and GIS. *J. Transp. Saf*, **2013**, 5(1), 27-45.
105. Borsati, M., Cascarano, M., & Bazzana, F. On the impact of average speed enforcement systems in reducing highway accidents: Evidence from the Italian Safety Tutor. *Econ. Transp*, **2019**, 20, 100123.
106. Dong, C., Dong, Q., Huang, B., Hu, W., & Nambisan, S. S. Estimating factors contributing to frequency and severity of large truck-involved crashes. *J. Transp. Eng. A: Sys*, **2017**, 143(8), 04017032.
107. Chang, G. L., & Xiang, H. (2003). The relationship between congestion levels and accidents (No. MD-03-SP 208B46,).
108. Yang, Y., & Zhang, Q. A hierarchical analysis for rock engineering using artificial neural networks. *Rock Mech Rock Eng*, **1997**, 30(4), 207-222.
109. Mikaeil, R., Mokhtarian, M., Haghshenas, S. S., Careddu, N., & Alipour, A. Assessing the System Vibration of Circular Sawing Machine in Carbonate Rock Sawing Process Using Experimental Study and Machine Learning. *Geotech. Geol. Eng*, **2021**, 1-17.
110. Lee, Y. M., Chong, S. Y., Goonting, K., & Sheppard, E. The effect of speed limit credibility on drivers' speed choice. *Transp. Res. F: Traffic Psychol. Behav.*, **2017**, 45, 43-53.
111. Vadeby, A., & Forsman, Å. Traffic safety effects of new speed limits in Sweden. *Accid Anal Prev*, **2018**, 114, 34-39.
112. Yu, R., & Abdel-Aty, M. Investigating the different characteristics of weekday and weekend crashes. *J Safety Res*, **2013**, 46, 91-97.
113. Mokhtarimousavi, S., Anderson, J. C., Azizinamini, A., & Hadi, M. Factors affecting injury severity in vehicle-pedestrian crashes: a day-of-week analysis using random parameter ordered response models and artificial neural networks. *Int. J. Transp. Sci. Technol*, **2020**, 9(2), 100-115.
114. Asgarzadeh, M., Fischer, D., Verma, S. K., Courtney, T. K., & Christiani, D. C. The impact of weather, road surface, time-of-day, and light conditions on severity of bicycle-motor vehicle crash injuries. *Am. J. Ind. Med*, **2018**, 61(7), 556-565.