Article

Feasibility of Stochastic Models for Evaluation of Potential Factors for Safety: A Case Study in Southern Italy

Giuseppe Guido ¹ , Sina Shaffiee Haghshenas ¹ , Sami Shaffiee Haghshenas ¹ ,
Alessandro Vitale ¹ , Vittorio Astarita ^{1,*} and Ashkan Shafiee Haghshenas ²

¹ Department of Civil Engineering, University of Calabria, Via Bucci, 87036 Rende, Italy; giuseppe.guido@unical.it (G.G.); s.shaffiee@yahoo.com (S.S.H.); sami.shaffiee@gmail.com (S.S.H.); alessandro.vitale@unical.it (A.V.)

² W Booth School of Engineering Practice & Technology, McMaster University, Main St W 1280, Hamilton, ON L8S 4L8, Canada; shafiea@mcmaster.ca

* Correspondence: vittorio.astarita@unical.it; Tel.: +39-098-449-6780

Received: 31 July 2020; Accepted: 11 September 2020; Published: 12 September 2020

Abstract: There is no definite conclusion about what the main variables that play a fundamental role in road safety are. Therefore, the identification of significant factors in road accidents has been a primary concern of the transportation safety research community for many years. Every accident is influenced by multiple variables that, in a given time interval, concur to cause a crash scenario. Information coming from crash reports is very useful in traffic safety research, and several reported crash variables can be analyzed with modern statistical methods to establish whether a classification or clustering of different crash variables is possible. Hence, this study aims to use stochastic techniques for evaluating the role of some variables in accidents with a clustering analysis. The variables that are considered in this paper are light conditions, weekday, average speed, annual average daily traffic, number of vehicles, and type of accident. For this purpose, a combination of particle swarm optimization (PSO) and the genetic algorithm (GA) with the k-means algorithm was used as the machine-learning technique to cluster and evaluate road safety data. According to a multiscale approach, based on a set of data from two years of crash records collected from rural and urban roads in the province of Cosenza, 154 accident cases were accurately investigated and selected for three categories of accident places, including straight, intersection, and other, in each urban and rural network. PSO had a superior performance, with 0.87% accuracy on urban and rural roads in comparison with GA, although the results of GA had an acceptable degree of accuracy. In addition, the results show that, on urban roads, social cost and type of accident had the most and least influence for all accident places, while, on rural roads, although the social cost was the most notable factor for all accident places, the type of accident had the least effect on the straight sections and curves, and the number of vehicles had the least influence at intersections.

Keywords: road safety; urban and rural networks; machine learning; particle swarm optimization (PSO); genetic algorithms (GA); stochastic techniques

1. Introduction

Traffic accidents are in the top 10 major causes of injuries and deaths [1], representing, with the associated economic loss, an important socioeconomic concern. For these reasons, the identification of the significant factors associated with such crashes has become of major interest and a great challenge in transportation safety research. There are many problems related to the inadequacy and difficulty of obtaining data for analysis as well as the limited crash counts. Moreover, another challenge with

such studies is that there may potentially be redundant information and correlations among different candidate factors. Consequently, considerable efforts have to be made in the field of accident analysis to address these problems properly and effectively.

Traditionally, in the scientific literature, statistical models based on crash frequency are a common tool for estimating the safety performance functions of many transportation systems. Count regression models are very often applied in road safety studies dealing with the crash-frequency context [2]. Among them there are Poisson or negative binomial (NB) models [3–7] and zero-inflated Poisson or NB models [5,8,9]. These safety models offer, as an advantage, the opportunity to correlate the specific roadway characteristics and expected crash frequencies. On the contrary, their limits are associated with strong parametric modeling assumptions and omitted variable bias, reducing the prediction suitability. In order to overcome the limits of traditional safety models, several researchers introduced heterogeneity models [10–12]. Random parameter models belong to this category of models [13–15]; in them, model parameters are allowed to vary from site to site, improving statistical inference and the predictive capability of such models [12]. In addition to traditional and heterogeneity safety models, researchers have proposed many nonparametric and artificial intelligence models for the study of crash injury patterns. Artificial intelligence approaches have been proven to be more efficient and effective in the modeling process [16]. These models are data driven and include decision trees (DT) [17–19], support vector machine (SVM) [20,21], and random forest (RF) [22]. The SVM has been utilized for the classification of crash injury severity [23,24] and outperforms the popular ordered probit model for the prediction of injury severity and factor impact assessment [23]. Thanks to the big-data era, optimization efficiency has become a primary concern for the analysis procedure. For these reasons, in the scientific literature, there is a need to develop novel algorithms to handle the traffic crash records efficiently.

The main aim of this paper is to propose a hybrid model combining particle swarm optimization (PSO) and genetic algorithm (GA) with k-means to cluster and to evaluate significant factors affecting road safety. The paper is organized as follows. A literature review of the different model categories and approaches is provided in the next section. Section 2 describes the methodological approach and the theoretical fundamentals of particle swarm optimization (PSO) and the genetic algorithm (GA). The case study and the data description are provided in Section 3. In Sections 4 and 5, the application of PSO and GA algorithms have been described. Sections 6 and 7 refer to the discussion of results and conclusions.

2. Methodology

To cluster a set of unpredicted factors for road safety evaluation, two different stochastic techniques were applied: particle swarm optimization (PSO) and genetic algorithm (GA) with k-means. In both cases, as in several data mining applications, the optimization problem considered is the minimization of Euclidean distance among the different groups of factors. The results obtained by applying the two different techniques were compared in terms of the performance and accuracy of the determined solutions. As a final result, the selected factors were used to create a basic road safety pattern with its temporal and spatial property and dimension.

2.1. Genetic Algorithm (GA)

In artificial intelligence, the genetic algorithm (GA), an evolutionary algorithm, is a powerful tool in the evaluation of complex systems. It was introduced and developed by John Holland and his colleagues during the two decades from 1960 [25]. GA works based upon a stochastic search technique and GA's performance is inspired by biological evolution. It has been widely applied in various fields of engineering, including computer sciences [26,27], transportation, and geotechniques [28–31] as well as mathematics [32]. There are three main operators in the process of GA, which include reproduction, crossover, and mutation rules. Generally speaking, at first, in the process of GA, the initial population is randomly generated as a set of number strings (representation of solutions). Then, the goodness and efficiency of each solution is evaluated by the fitness function (objective function),

and the best population (solutions) is selected as generators (parents) based upon a reproduction operator to generate new generations (new solutions) based upon the crossover operator. According to the mutation operator, a random change emerges in the next generation, whereby this new generation has higher fitness values in comparison with the old generation. This loop continues and the accuracy of the process is investigated; if it reaches a desired optimization, the loop is stopped [33]. The basic form of the flowchart of GA is shown in Figure 1.

Figure 1. The basic form of the flowchart of the genetic algorithm GA [34].

2.2. Particle Swarm Optimization (PSO)

In the modern world, most phenomena and problems deal with complex and nonlinear systems; therefore, they cannot be easily solved. Hence, using artificial intelligence techniques is imperative [35–41]. Metaheuristic algorithms, a branch of artificial intelligence techniques, are the precise scientific tools to solve uncertain systems in a wide range of sciences and industries [42–49]. Particle swarm optimization (PSO) is one of the notable metaheuristic algorithms that is appropriate for dealing with nonlinear relationships and complex problems. Recently, it has been used successfully to solve complex engineering problems determined by many parameters [28,50,51]. PSO was introduced by Kennedy [52,53]. PSO is inspired by swarm intelligence (SI) and the social behavior of birds and fish in finding enough food in a group. In the process of PSO, a set of solutions, which is a set of particles, is randomly generated; in fact, each solution is a particle. This algorithm works in a D-dimensional community or search space and each particle can affect other neighboring particles or the general particles. Each particle has a velocity vector and, for determining the best optimal answer (new position), two factors of personal best position (Pbest) and global best position (Gbest) are considered [54,55]. Figure 2 shows the movement of particles and their new positions, and the updated velocity vector and position are shown in Equations (1) and (2).

$$V_i^{(k+1)} = wV_i^k + c_1r_1 \cdot (pbest_i - X_i^k) + c_2r_2 \cdot (gbest - X_i^k) \quad (1)$$

$$X_i^{(k+1)} = X_i^k + V_i^k \quad (2)$$

where $i = (1, 2, 3, \dots, N)$ and N is population size (particle). $V_i^{(k+1)}$ and $V_i^{(k)}$ are the new velocity vector and old velocity vector of the i th particle, respectively. $X_i^{(k+1)}$ and $X_i^{(k)}$ represent the new position and old position of the i th particle. Furthermore, $K = (1, 2, 3, \dots)$ is the number of iterations. w is the inertia weight that is in the interval of 0.4 to 0.9, and it plays a key role in controlling the effect of the old velocity vector on the new velocity vector; r_1 and r_2 are random numbers that are in the interval of 0 and +1. C_1 and C_2 are the acceleration constants that should be evaluated by experts and based upon Equation (3). In fact, C_1 and C_2 are known as the individual learning and social learning factors, respectively [56,57]. Figure 3 shows the movement and modified position of a particle. The particles continue to search for the best position (solution) and, when they find the best position, the algorithm process is stopped. It is worth mentioning that PSO can be applied to solve continuous and discrete problems.

$$c_1 + c_2 \leq 4 \quad (3)$$

Figure 2. New velocity and position vectors of a set of particles in a search space.

Figure 3. Concept of the modification of the movement of a particle by the particle swarm optimization (PSO) algorithm [58].

2.3. The Optimization Function and the Correlation Analysis Method

As mentioned before, the fitness function contributes greatly to the determination of the solution. Hence, Lloyd's algorithm (k-means clustering) is considered the fitness function for two metaheuristic

algorithms according to the problem statement in this research work. Equation (4) represents Lloyd's algorithm, which is a common and popular clustering algorithm in data mining [59].

$$Obj.Function = \sum_{i=1}^n \min_{1 \leq j \leq k} d(x_i, m_j) \quad (4)$$

where x_i and m_j are the data set and center of each cluster, respectively, and $j = (1, 2, 3, \dots, k)$ indicates the number of the cluster. Furthermore, d is the Euclidean distance between each member and the center of each cluster. It is worth mentioning that two aims are considered (for the Lloyd's algorithm) the fitness function: these are the distance of the centers of the clusters from each other and also the distance between the members of a cluster, which should become maximization and minimization, respectively.

Before modeling, correlation analysis of the data set is imperative because a high correlation between data sets can have a negative impact on the performances and results of clustering algorithms. Hence, although the dataset was considered by experts in this study, a correlation analysis was conducted. Pearson's correlation coefficient, a suitable and common correlation analysis, was used for the evaluation of the data set. Mathematical relations are shown in Equations (5) to (8), in which $\rho(r)$ is known as Pearson's correlation coefficient for two independent parameters (X and Y). SP_{Dxy} is the covariance of X and Y. SS_X and SS_Y indicate the standard deviation of X and Y, respectively [60–62]. It is worth mentioning that the value of $\rho(r)$ is in the interval of -1 and $+1$, in which the absolute values of the coefficient demonstrate degrees of correlation and the signs only represent a direct (+) and reverse (−) relation between two independent parameters.

$$\rho = r = \frac{SP_{Dxy}}{\sqrt{SS_X \cdot SS_Y}} \quad (5)$$

$$SP_{Dxy} = \sum xy - \frac{(\sum x)(\sum y)}{n} \quad (6)$$

$$SS_X = \sum_i x_i^2 - \frac{(\sum x_i)^2}{n} \quad (7)$$

$$SS_Y = \sum_i y_i^2 - \frac{(\sum y_i)^2}{n} \quad (8)$$

3. Data Collection and Preparation

3.1. Crash, Traffic, and Speed Data

The crash data used in this research have been derived from the large database of Istituto Nazionale di Statistica (ISTAT). Each crash record was obtained from the collaboration of a multiplicity of entities: ISTAT, Automobile Club Italia (ACI), the Ministry of the Interior, the traffic police department, the Carabinieri, the provincial police, the municipal police, the offices of statistics of the provincial capitals, and the statistical offices of some provinces that have signed an agreement with ISTAT. The database contains details on road accidents that were reported by a police authority, occurred in the national territory over a calendar year, and caused injuries to people (death or injury). There is information on crash dynamics and location, on vehicles, and on the individuals involved, but it does not include property damage only (PDO) events in accordance with current Italian legislation. Indeed, Italian legislation defines road accidents as crashes only when they result in at least one injury [63].

In this research, the road accidents that were selected occurred during two years (2017 and 2018) on the main rural and urban roads of the province of Cosenza (Figure 4). A total of 154 crash events were elaborated, considering 77 accidents on rural roads and 77 on urban roads. In order to analyze and compare homogeneous road accident samples by applying PSO and GA techniques, two samples

of 77 crashes for both urban and rural cases were extracted from an accident database. These crashes have been distinguished and grouped by territorial area (urban and rural) and homogeneous geometric elements (intersection, straight, curve, other, etc.). The ISTAT database fields considered in this study are summarized in Table 1.

Figure 4. Map of Cosenza province with accident locations (source: CRISC, Calabria).

Table 1. Istituto Nazionale di Statistica (ISTAT) database fields considered.

Data Field Type	Data Field	Description
Human characteristic	Driver gender	Male or female
Vehicle characteristic	Vehicle type	Car, motorcycle, truck, and other
Road environment	Road type	National rural road, provincial rural road, national and provincial rural road in urban context, and urban road
Other environment	Light conditions	Daylight and nighttime
	Day of the week	Weekday and weekend
Location environment	Macroarea location	Urban and rural
	Number of vehicles	Number of vehicles involved
Accident characteristic	Accident nature	Way out, collision with an accidental obstacle, side collision, front-side collision, rear-end collision, head-on collision, pedestrian collision, impact with parked vehicle, impact with stopped vehicle, fall from vehicle, and sudden braking
	Accident severity	Injuries and deaths

The fields of the ISTAT database were matched up with ad hoc traffic surveys on the same rural and urban roads considered in order to derive average vehicular speeds and average traffic volumes. The surveys were conducted for one month in October 2019 with the support of Bluetooth radar sensors, which give output not only on traffic volumes but also vehicular speed values. In particular, radar sensors were located on the road sections with observed crashes. From an evaluation of traffic-volume and speed-value statistical trends over a 10-year period, and considering the social,

economic, demographic, and travel demand characteristics of the study area, traffic volumes and vehicular speed values were considered invariant over the last five years. An important aspect for the present research was the definition of a segment surrounding each count station, where it could be assumed that homogeneous flow and speed conditions were present for the entire length. If the sensor was positioned on a link with homogeneous geometric characteristics greater than 2 miles in length, a circular buffer 2 miles in diameter around the location of the radar sensor (1 mile upstream and downstream) was defined [64]. The geometric homogeneity of a road segment was defined, taking into account the number of lanes, lane and shoulder width, speed limit, median type, and median width.

3.2. Correlation Data Analysis

Six unpredicted and effective variables in accidents, including the daylight, weekday, average speed, annual average daily traffic, number of vehicles, and type of accident were considered for urban and rural roads. As described in Section 2.3, during the first step, 77 accident cases for urban and rural roads related to the six mentioned variables were normalized between 0 and 1 and then correlation analyses were conducted.

The results of the correlation analysis for urban and rural roads are shown in Tables 2 and 3.

Table 2. Pearson's correlation coefficient for the urban area dataset.

	Daylight	Weekday	Average Speed	Annual Average Daily Traffic	Number of Vehicles	Type of Accident
Daylight	1					
Weekday	−0.309	1				
Average speed	−0.124	0.262	1			
Annual average daily traffic	0.042	−0.246	−0.378	1		
Number of vehicles	−0.012	−0.112	0.148	0.104	1	
Type of accident	−0.035	−0.029	0.06	0.028	0.182	1

Table 3. Pearson's correlation coefficient for rural area dataset.

	Daylight	Weekday	Average Speed	Annual Average Daily Traffic	Number of Vehicles	Type of Accident
Daylight	1					
Weekday	−0.327	1				
Average speed	−0.078	0.118	1			
Annual average daily traffic	0.158	0.109	0.363	1		
Number of vehicles	−0.049	0.215	0.09	0.053	1	
Type of accident	−0.204	0.279	−0.08	0.023	0.396	1

It should be noted that when $|p| > 0.85$ is established, the correlation between two parameters was considered strong and was not appropriate for clustering analysis. According to the results of Tables 3 and 4, the maximum correlation was -0.378 between the average speed and annual average daily traffic in the urban areas; they have a reverse relation with each other. On the other hand, there was the maximum correlation between the number of vehicles and type of accident of 0.396 , with a direct relation between the two in rural areas. Hence, it can be concluded that all parameters in this study have a suitable correlation with each other, which is in full agreement with other studies on road safety [63].

4. Particle Swarm Optimization Modeling for Urban and Rural Area

Before modeling, some control parameters of the PSO algorithm should be determined at the first step. In order to reach a high level of precision in the performance of the algorithm, it is necessary to

determine suitable control parameters, which contribute greatly to the increase of the convergence of the algorithm. There are no determined relations; in fact, these parameters were determined based upon trial and error and experts' opinions [65,66]. The personal learning coefficient (C_1) and global learning coefficient (C_2) were considered equal to 2. The minimum acceptance precision, $\varepsilon_L = 0.00001$, and the maximum number of iterations and population size (swarm size) were considered 300 and 50, respectively. It is worth mentioning that these control parameters can be considered the same for urban and rural situations because the number and type of dataset are the same. Then, modeling was carried out by three classes (clusters) in a Matlab software environment. Figure 5 shows the best cost of performance of the model of the PSO algorithm for urban and rural areas.

Figure 5. The performance of the PSO algorithm based upon the best cost of fitness function for urban and rural areas.

According to Figure 5, the convergence trends of the algorithm for both areas shows that the performance of the algorithm was reliable because, after the 76th and 39th iterations, the convergence trends remained constant until the 300th iteration for urban and rural areas, respectively. Optimization and clustering of accident cases by the PSO algorithm for urban and rural areas are shown in Tables 4 and 5, respectively.

According to the obtained results in Tables 4 and 5, the Euclidean distance of each accident case from each class (cluster) was calculated based upon the six unpredicted and effective variables in accidents and the lowest distance shows which accident case belongs to which class, and the results are shown in the column of the recognized class. For example, the minimum Euclidean distance in the second case in urban areas was 0.05 from the first class, which shows that this accident case belongs to the first class. In rural areas, the 47th accident case with the minimum Euclidean distance of 0.899 from the third class shows that this case belongs to the third class. Then, the results of the column of the recognized class were compared with the column for the actual type of accident place. Ten cases were wrongly recognized from 77 accident cases for both areas. For instance, in urban areas, the minimum Euclidean distance of the 40th case was calculated as 0.021, which belongs to the first class, although this accident case occurred at the intersection in the actual type. In addition, the 70th case should be recognized at the intersection (second class), although, due to the minimum Euclidean distance, it was recognized in the third class. Generally, the accuracy of the algorithm's performance was 0.87% for both urban and rural networks, and it should be noted that clustering of accident cases for both areas by model had an acceptable degree. The impacts of six unpredicted and effective variables on each category of accident places based upon the Euclidean distance are shown in Table 6.

Table 4. Optimization and clustering of accident cases in urban area by PSO.

Case #	Optimum Partition			Rec. Class	Actual Type of Accident Place	Case #	Optimum Partition			Rec. Class	Actual Type of Accident Place
	The First Class	The Second Class	The Third Class				The First Class	The Second Class	The Third Class		
1	0.327	1.051	1.033	1	Straight	40	0.021	1.014	1.016	1	Intersection
2	0.050	1.017	1.016	1	Straight	41	0.107	1.020	1.014	1	Straight
3	1.131	1.498	0.405	3	Other	42	0.471	1.096	1.114	1	Straight
4	1.472	0.937	1.692	2	Intersection	43	0.227	1.054	1.063	1	Straight
5	1.424	0.869	1.664	2	Intersection	44	0.997	1.321	1.398	1	Straight
6	0.278	1.072	1.088	1	Straight	45	0.021	1.014	1.016	1	Straight
7	1.023	1.436	0.189	3	Other	46	0.389	1.101	1.125	1	Straight
8	0.050	1.017	1.016	1	Straight	47	0.236	1.044	1.048	1	Straight
9	0.050	1.017	1.016	1	Straight	48	1.273	1.539	1.651	1	Straight
10	0.358	1.065	1.043	1	Straight	49	0.255	1.043	1.026	1	Straight
11	0.766	1.238	1.259	1	Straight	50	0.471	1.096	1.114	1	Straight
12	1.001	0.185	1.359	2	Intersection	51	1.414	0.921	1.014	2	Intersection
13	0.278	1.072	1.088	1	Straight	52	1.022	0.235	1.360	2	Straight
14	0.317	1.068	1.063	1	Straight	53	1.083	1.470	0.301	3	Straight
15	1.485	0.952	1.727	2	Intersection	54	1.497	0.979	1.749	2	Intersection
16	0.217	1.030	1.018	1	Straight	55	1.105	1.484	0.491	3	Other
17	1.603	1.112	1.823	2	Intersection	56	1.010	0.193	1.355	2	Intersection
18	0.205	1.047	1.052	1	Straight	57	1.586	1.792	1.224	3	Straight
19	1.001	0.185	1.359	2	Intersection	58	1.023	0.234	1.359	2	Straight
20	0.050	1.017	1.016	1	Straight	59	1.400	1.113	1.498	2	Intersection
21	0.298	1.052	1.033	1	Straight	60	0.205	1.047	1.052	1	Intersection
22	0.205	1.047	1.052	1	Straight	61	0.471	1.096	1.114	1	Intersection
23	1.010	1.337	1.429	1	Straight	62	1.055	1.079	0.410	3	Other

Table 4. Cont.

Case #	Optimum Partition			Rec. Class	Actual Type of Accident Place	Case #	Optimum Partition			Rec. Class	Actual Type of Accident Place
	The First Class	The Second Class	The Third Class				The First Class	The Second Class	The Third Class		
24	0.217	1.029	1.022	1	Straight	63	1.481	1.132	1.075	3	Straight
25	0.051	1.016	1.020	1	Straight	64	0.205	1.047	1.052	1	Straight
26	0.392	1.103	1.129	1	Straight	65	0.205	1.047	1.052	1	Straight
27	0.021	1.014	1.016	1	Straight	66	0.019	1.014	1.014	1	Straight
28	0.698	1.217	1.252	1	Straight	67	0.471	1.096	1.114	1	Straight
29	0.236	1.034	1.016	1	Straight	68	1.421	0.886	1.687	2	Intersection
30	1.006	1.428	0.169	3	Other	69	0.189	1.039	1.052	1	Straight
31	0.723	1.238	1.283	1	Straight	70	1.220	0.693	1.544	2	Intersection
32	0.403	1.107	1.123	1	Straight	71	0.507	1.120	1.148	1	Straight
33	0.516	1.109	1.115	1	Straight	72	0.205	1.047	1.052	1	Straight
34	0.471	1.096	1.114	1	Straight	73	0.255	1.043	1.026	1	Straight
35	0.107	1.020	1.014	1	Straight	74	0.050	1.012	1.013	1	Straight
36	0.021	1.014	1.016	1	Straight	75	0.205	1.047	1.052	1	Straight
37	0.166	1.000	1.359	1	Straight	76	0.050	1.012	1.013	1	Straight
38	1.019	1.331	1.399	1	Intersection	77	1.023	0.226	1.358	2	Intersection
39	0.471	1.096	1.114	1	Intersection						

Table 5. Optimization and clustering of accident cases in rural area by PSO.

Case #	Optimum Partition			Rec. Class	Actual Type of Accident Place	Case #	Optimum Partition			Rec. Class	Actual Type of Accident Place
	The First Class	The Second Class	The Third Class				The First Class	The Second Class	The Third Class		
1	0.132	1.011	1.151	1	Straight	40	1.403	0.964	1.316	2	Intersection
2	1.405	1.711	0.727	3	Other	41	1.134	1.501	0.728	3	Other
3	0.152	1.011	1.152	1	Straight	42	0.030	1.007	1.153	1	Straight
4	0.132	1.011	1.151	1	Straight	43	0.073	1.012	1.156	1	Straight
5	1.422	1.725	0.775	3	Other	44	0.235	1.030	1.184	1	Straight
6	0.273	1.045	1.196	1	Straight	45	0.235	1.030	1.184	1	Straight
7	1.436	1.029	0.781	3	Other	46	0.192	1.027	1.163	1	Straight
8	1.009	1.422	0.570	3	Other	47	1.740	1.405	0.899	3	Other
9	1.416	1.723	0.742	3	Other	48	1.011	0.183	1.241	2	Straight
10	0.128	1.019	1.240	1	Straight	49	0.244	1.035	1.188	1	Straight
11	0.128	1.019	1.240	1	Straight	50	1.030	0.244	1.269	2	Straight
12	0.116	1.000	1.239	1	Straight	51	0.126	1.014	1.154	1	Straight
13	1.392	0.995	1.235	2	Intersection	52	0.273	1.045	1.196	1	Straight
14	0.235	1.030	1.184	1	Straight	53	1.088	1.466	0.664	3	Other
15	0.235	1.030	1.184	1	Straight	54	1.036	0.215	1.249	2	Straight
16	1.424	1.728	0.781	3	Other	55	1.394	0.992	1.237	2	Intersection
17	1.728	1.392	0.854	3	Other	56	1.389	0.984	1.236	2	Intersection
18	0.235	1.030	1.184	1	Straight	57	1.436	1.029	0.781	3	Other
19	1.012	1.422	0.572	3	Other	58	0.235	1.030	1.184	1	Straight
20	0.082	1.006	1.153	1	Straight	59	0.235	1.030	1.184	1	Straight
21	0.132	1.011	1.151	1	Straight	60	0.132	1.011	1.151	1	Straight
22	0.152	1.011	1.152	1	Straight	61	0.235	1.030	1.184	1	Straight
23	1.009	1.420	0.569	3	Other	62	1.436	1.029	0.781	3	Other

Table 5. Cont.

Case #	Optimum Partition			Rec. Class	Actual Type of Accident Place	Case #	Optimum Partition			Rec. Class	Actual Type of Accident Place
	The First Class	The Second Class	The Third Class				The First Class	The Second Class	The Third Class		
24	0.030	1.007	1.153	1	Straight	63	1.030	0.244	1.269	2	Straight
25	0.030	1.007	1.153	1	Straight	64	0.986	1.393	1.239	1	Intersection
26	0.384	1.057	1.188	1	Other	65	0.030	1.007	1.153	1	Straight
27	1.728	1.390	0.854	3	Intersection	66	0.321	1.046	1.175	1	Straight
28	0.992	1.392	1.234	1	Other	67	0.261	1.023	1.163	1	Straight
29	0.235	1.030	1.184	1	Straight	68	0.177	1.024	1.160	1	Straight
30	0.082	1.006	1.153	1	Straight	69	1.093	1.470	0.670	3	Other
31	0.428	1.073	1.200	1	Straight	70	1.425	1.024	0.742	3	Intersection
32	1.030	0.244	1.269	2	Straight	71	0.030	1.007	1.153	1	Straight
33	1.728	1.390	0.854	3	Other	72	0.177	1.024	1.160	1	Straight
34	0.235	1.030	1.184	1	Straight	73	0.315	1.060	1.208	1	Straight
35	0.131	1.008	1.151	1	Straight	74	1.737	1.398	0.863	3	Other
36	0.244	1.035	1.188	1	Straight	75	0.382	1.067	1.191	1	Straight
37	1.409	0.968	1.315	2	Intersection	76	1.398	1.009	1.240	2	Intersection
38	0.235	1.030	1.184	1	Straight	77	0.177	1.024	1.160	1	Straight
39	1.000	1.419	0.573	3	Other						

Table 6. Euclidean distance of each effective variable from the center of each class in urban and rural areas for PSO.

Urban Area						
	Daylight	Weekday	Average Speed (km/h)	Annual Average Daily Traffic (Vehicle/Day)	Number of Vehicles	Type of Accident
Straight	0.000	0.991	0.799	0.309	0.411	1.000
Intersection	1.000	0.852	0.732	0.324	0.369	1.000
Other	0.092	0.963	0.703	0.258	0.308	0.000
Rural Area						
	Daylight	Weekday	Average Speed (km/h)	Annual Average Daily Traffic (Vehicle /Day)	Number of Vehicles	Type of Accident
Straight	0.016	1.000	0.769	0.283	0.147	1.000
Intersection	0.042	1.000	0.765	0.208	0.103	0.000
Other	0.401	0.000	0.711	0.229	0.119	0.603

According to the obtained results, daylight had the most influence on accident cases on straight roads for urban and rural areas (0 and 0.016, respectively) because, in both of them, daylight had the lowest Euclidean distance from straight in comparison with other effective variables. On the other hand, annual average daily traffic plays a key role in accident cases at intersections in urban areas, with the lowest Euclidean distance of 0.324, while the type of accident was the most important variable of accident cases at the intersection, with the lowest Euclidean distance of 0 among six variables in rural areas. Finally, the type of accident and the day of the week had the most impact on accident cases in the third category of accident places (other) in comparison with other variables with the lowest Euclidean distance of 0 in urban and rural areas.

5. Genetic Algorithm Modeling for Urban and Rural Areas

As mentioned before, the control parameters of the algorithm play a key role in the faster convergence of the algorithm. Hence, according to experts' experiences and the number of data sets in this study, some of these parameters were selected, namely, the minimum acceptance precision, $\varepsilon_L = 0.00001$, population size (PS) = 250, maximum number of iterations = 250, and crossover and mutation rate (MR), which were 0.8 and 0.02, respectively [67,68]. It should be noted that these values are dimensionless. After determining the control parameters, the models were constructed with a pseudocode of GA in a Matlab environment. With regards to the data set and problem statement in this research work, the number of classes was considered to be three (clusters) for the urban and rural areas dataset. The performance of models according to the best cost is demonstrated in Figure 6.

Figure 6. The performance of the PSO algorithm based upon the best cost of fitness function for urban and rural areas.

According to Figure 6, the convergence trends achieved the desired precision level and it was fixed from the 94th to the 300th iteration for the urban dataset, and the convergence trend was fixed from the 102nd to the 300th iteration for the rural dataset. Hence, it can be concluded that the convergence trends of the algorithm had acceptable degrees and speed for urban and rural situations. Tables 7 and 8 show the optimization and clustering of accident cases by the GA algorithm for the urban and rural areas, respectively.

Table 7. Optimization and clustering of accident cases in urban area by GA.

Case #	Optimum Partition			Rec. Class	Actual Type of Accident Place	Case #	Optimum Partition			Rec. Class	Actual Type of Accident Place
	The First Class	The Second Class	The Third Class				The First Class	The Second Class	The Third Class		
1	0.327	1.051	1.033	1	Straight	40	0.021	1.014	1.016	1	Intersection
2	0.050	1.018	1.016	1	Straight	41	0.107	1.021	1.014	1	Straight
3	1.131	1.498	0.406	3	Other	42	0.471	1.096	1.113	1	Straight
4	1.472	0.934	1.689	2	Intersection	43	0.227	1.054	1.063	1	Straight
5	1.424	0.867	1.661	2	Intersection	44	0.996	1.320	1.397	1	Straight
6	0.278	1.073	1.088	1	Straight	45	0.021	1.014	1.016	1	Straight
7	1.023	1.436	0.193	3	Other	46	0.389	1.101	1.124	1	Straight
8	0.050	1.018	1.016	1	Straight	47	0.236	1.044	1.047	1	Straight
9	0.050	1.018	1.016	1	Straight	48	1.273	1.537	1.649	1	Straight
10	0.358	1.066	1.043	1	Straight	49	0.255	1.044	1.027	1	Straight
11	0.766	1.236	1.257	1	Straight	50	0.471	1.096	1.113	1	Straight
12	1.001	0.188	1.357	2	Intersection	51	1.414	1.014	0.917	3	Intersection
13	0.278	1.073	1.088	1	Straight	52	1.022	0.238	1.358	2	Straight
14	0.317	1.068	1.063	1	Straight	53	1.083	1.471	0.303	3	Straight
15	1.484	0.949	1.723	2	Intersection	54	1.496	0.976	1.745	2	Intersection
16	0.217	1.031	1.018	1	Straight	55	1.105	1.483	0.489	3	Other
17	1.603	1.108	1.820	2	Intersection	56	1.010	0.196	1.352	2	Intersection
18	0.205	1.048	1.052	1	Straight	57	1.585	1.790	1.222	3	Straight
19	1.001	0.188	1.357	2	Intersection	58	1.023	0.237	1.356	2	Straight
20	0.050	1.018	1.016	1	Straight	59	1.113	1.397	1.496	1	Intersection
21	0.298	1.053	1.033	1	Straight	60	0.205	1.048	1.052	1	Intersection
22	0.205	1.048	1.052	1	Straight	61	0.471	1.096	1.113	1	Intersection
23	1.010	1.336	1.428	1	Straight	62	0.410	1.079	1.054	1	Other

Table 7. Cont.

Case #	Optimum Partition			Rec. Class	Actual Type of Accident Place	Case #	Optimum Partition			Rec. Class	Actual Type of Accident Place
	The First Class	The Second Class	The Third Class				The First Class	The Second Class	The Third Class		
24	0.217	1.030	1.022	1	Straight	63	1.481	1.132	1.072	3	Straight
25	0.051	1.016	1.020	1	Straight	64	0.205	1.048	1.052	1	Straight
26	0.392	1.103	1.128	1	Straight	65	0.205	1.048	1.052	1	Straight
27	0.021	1.014	1.016	1	Straight	66	0.020	1.015	1.014	1	Straight
28	0.698	1.216	1.251	1	Straight	67	0.471	1.096	1.113	1	Straight
29	0.237	1.034	1.016	1	Straight	68	1.420	0.883	1.684	2	Intersection
30	1.006	1.429	0.172	3	Other	69	0.189	1.040	1.051	1	Straight
31	0.723	1.237	1.281	1	Straight	70	1.219	0.692	1.541	2	Intersection
32	0.403	1.107	1.123	1	Straight	71	0.507	1.119	1.147	1	Straight
33	0.516	1.109	1.114	1	Straight	72	0.205	1.048	1.052	1	Straight
34	0.471	1.096	1.113	1	Straight	73	0.255	1.044	1.027	1	Straight
35	0.107	1.021	1.014	1	Straight	74	0.050	1.012	1.013	1	Straight
36	0.021	1.014	1.016	1	Straight	75	0.205	1.048	1.052	1	Straight
37	1.000	0.168	1.356	2	Straight	76	0.050	1.012	1.013	1	Straight
38	1.019	1.330	1.398	1	Intersection	77	1.023	0.229	1.356	2	Intersection
39	0.471	1.096	1.113	1	Intersection						

Table 8. Optimization and clustering of accident cases in rural area by GA.

Case #	Optimum Partition			Recognized Class	Actual Type of Accident Place	Case #	Optimum Partition			Recognized Class	Actual Type of Accident Place
	The First Class	The Second Class	The Third Class				The First Class	The Second Class	The Third Class		
1	0.133	1.019	1.054	1	Straight	40	1.414	0.964	1.155	2	Intersection
2	1.416	1.654	0.441	3	Other	41	1.136	1.451	0.890	3	Other
3	0.153	1.019	1.056	1	Straight	42	0.021	1.010	1.058	1	Straight
4	0.133	1.019	1.054	1	Straight	43	0.071	1.016	1.061	1	Straight
5	1.433	1.665	0.523	3	Other	44	0.233	1.027	1.096	1	Straight
6	0.272	1.043	1.107	1	Straight	45	0.233	1.027	1.096	1	Straight
7	1.435	0.929	1.148	2	Other	46	0.193	1.035	1.065	1	Straight
8	1.009	1.359	0.772	3	Other	47	1.749	1.331	0.952	3	Other
9	1.427	1.670	0.453	3	Other	48	1.011	0.225	1.321	2	Straight
10	1.019	0.185	1.324	2	Straight	49	0.243	1.032	1.099	1	Straight
11	1.019	0.185	1.324	2	Straight	50	1.030	0.234	1.355	2	Straight
12	1.000	0.146	1.322	2	Straight	51	0.128	1.021	1.056	1	Straight
13	1.012	1.395	0.839	3	Intersection	52	0.272	1.043	1.107	1	Straight
14	0.233	1.027	1.096	1	Straight	53	1.089	1.412	0.841	3	Other
15	0.233	1.027	1.096	1	Straight	54	1.037	0.265	1.331	2	Straight
16	1.435	1.668	0.529	3	Other	55	1.008	1.397	0.839	3	Intersection
17	1.737	1.325	0.901	3	Other	56	1.000	1.389	0.840	3	Intersection
18	0.233	1.027	1.096	1	Straight	57	1.435	0.929	1.148	2	Other
19	1.012	1.359	0.775	3	Other	58	0.233	1.027	1.096	1	Straight
20	0.078	1.010	1.060	1	Straight	59	0.233	1.027	1.096	1	Straight
21	0.133	1.019	1.054	1	Straight	60	0.133	1.019	1.054	1	Straight
22	0.153	1.019	1.056	1	Straight	61	0.233	1.027	1.096	1	Straight
23	1.009	1.357	0.773	3	Other	62	1.435	1.148	0.929	3	Other

Table 8. Cont.

Case #	Optimum Partition			Recognized Class	Actual Type of Accident Place	Case #	Optimum Partition			Recognized Class	Actual Type of Accident Place
	The First Class	The Second Class	The Third Class				The First Class	The Second Class	The Third Class		
24	0.021	1.010	1.058	1	Straight	63	1.030	0.234	1.355	2	Straight
25	0.021	1.010	1.058	1	Straight	64	1.002	0.844	1.393	2	Intersection
26	0.388	1.072	1.094	1	Other	65	0.021	1.010	1.058	1	Straight
27	1.737	1.322	0.902	3	Intersection	66	0.324	1.060	1.079	1	Straight
28	1.009	1.395	0.836	3	Other	67	0.265	1.035	1.068	1	Straight
29	0.233	1.027	1.096	1	Straight	68	0.176	1.032	1.062	1	Straight
30	0.078	1.010	1.060	1	Straight	69	1.094	1.416	0.846	3	Other
31	0.431	1.089	1.106	1	Straight	70	1.425	0.935	1.114	2	Intersection
32	1.030	0.234	1.355	2	Straight	71	0.021	1.010	1.058	1	Straight
33	1.737	1.322	0.902	3	Other	72	0.176	1.032	1.062	1	Straight
34	0.233	1.027	1.096	1	Straight	73	0.315	1.058	1.118	1	Straight
35	0.132	1.016	1.056	1	Straight	74	1.746	1.333	0.910	3	Other
36	0.243	1.032	1.099	1	Straight	75	0.385	1.082	1.095	1	Straight
37	1.420	0.973	1.152	2	Intersection	76	1.025	0.845	0.929	2	Intersection
38	0.233	1.027	1.096	1	Straight	77	0.176	1.032	1.062	1	Straight
39	1.000	1.353	0.776	3	Other						

The optimum partitions of three classes (clusters) were calculated for the urban and rural areas dataset, with these values being based upon the Euclidean distance and being dimensionless. For instance, in the first accident case of the urban area, the minimum Euclidean distance was 0.327, which belongs to the first class, and the results were in full agreement with the actual type of accident place. However, for the 37th accident case of the urban area, the minimum of Euclidean distance belonged to the second class, with 0.168 among three classes, so it did not match with the real case. In the rural area, the model could classify the second accident case in the third class, so it did match with the actual type of accident place. In addition, in the 63rd case of a rural area, the minimum Euclidean distance was 0.234 for the second class among the three classes that the model had an error for in this case because it was “straight” in the real case. Generally, the model of the GA algorithm can classify the urban and rural areas dataset with 82% and 81% accuracies, respectively. Table 9 shows the values of the Euclidean distance of each effective variable from the center of each class by GA.

Table 9. Euclidean distance of each effective variable from the center of each class in urban and rural areas for GA.

Urban Area						
	Daylight	Weekday	Average Speed (km/h)	Annual Average Daily Traffic (Vehicle/Day)	Number of Vehicles	Type of Accident
Straight	0.000	0.991	0.799	0.309	0.411	1.000
Intersection	1.000	0.850	0.730	0.326	0.370	0.999
Other	0.096	0.961	0.702	0.260	0.309	0.001
Rural Area						
	Daylight	Weekday	Average Speed (km/h)	Annual Average Daily Traffic (Vehicle /Day)	Number of Vehicles	Type of Accident
Straight	0.000	1.000	0.771	0.287	0.146	1.000
Intersection	0.046	0.904	0.795	0.220	0.101	0.000
Other	0.706	0.242	0.693	0.239	0.136	0.814

The obtained results in Table 9 show that daylight had the most impact for “straight” in the urban and rural areas in comparison with other variables because it had the lowest Euclidean distance equal to zero. The annual average daily traffic and type of accident were the most important variables for the intersection, with 0.326 and 0 in urban and rural areas, respectively. Finally, the type of accident was the most effective factor, with a Euclidean distance of 0.001 for other sections in the urban area, while the number of vehicles contributes greatly for other sections in rural areas in comparison with other variables.

6. Results and Discussion

As mentioned before, the aim of the present research work is to propose the feasibility of stochastic models for the evaluation of potential factors for safety. Owing to the ambiguity and uncertainty present in the investigative process of road safety, two stochastic techniques have been proposed for clustering and evaluating safety. Using a data mining approach, two models were developed for evaluating potential safety factors, using 77 accident cases in urban and rural areas in Cosenza in southern Italy. A comparison was made between the obtained results of two algorithms’ performances, based upon their accuracies, in Figure 7. Consequently, it was found that, although the PSO algorithm provided more accurate clustering in comparison with GA, both stochastic techniques were reliable system modeling techniques for clustering and evaluating potential safety factors.

Figure 7. Comparison of measured accuracies of two algorithms' performances for urban and rural areas.

Furthermore, the two stochastic techniques applied to the case study underlined several common issues.

Models of both algorithms recognized that daylight had the most impact on the “straight” in urban and rural areas and that these results were in full agreement with real conditions of the straight place in two areas of Cosenza.

Annual average daily traffic was considered the most important factor for the “intersection” place only in urban areas by models of PSO and GA. However, these models considered that the type of accident was the most effective variable in rural areas.

Furthermore, in comparison to the Euclidean distances of six variables, both algorithms recognized that the type of accident was the most effective variable for “other” in urban areas.

The unique difference between the two approaches is that, in rural areas, the PSO approach considered the weekday as the most important variable for “other”, while the number of vehicles was recognized by the GA model. The reason could be that both PSO and GA start from nearby solutions but, then, PSO converges very quickly while GA struggles up to a great number of iterations and then converges to a solution nearer to PSO, but this solution is still not better than the one provided by PSO.

It should be noted that these two models based upon the determined control parameters and obtained results are case-specific and cannot be used directly for evaluating the safety of transportation systems of other cities. In addition, these techniques are not appropriate for the evaluation of incomplete data.

7. Conclusions

This paper presented two stochastic techniques to cluster and evaluate transportation safety. For this purpose, a combination of PSO and the genetic algorithm (GA) with the k-means algorithm were used as well as six notable unpredicted variables in accidents, including the daylight, weekday, average speed, annual average daily traffic, number of vehicles, and type of accident, which were considered for urban and rural areas in Cosenza in southern Italy. Some 77 cases of accidents for urban and rural areas were selected and all the cases were classified into three classes. The obtained results were validated based upon three types of accident places, including straight, curve, and intersection, in each urban and rural network. Although both algorithms had a suitable capability to classify case accidents, the PSO algorithm had a superior performance, with 0.86% and 0.87% accuracies on urban and rural roads, respectively, in comparison with GA. In addition, the obtained results of the PSO algorithm show that daylight was the most important parameter to evaluate safety for the straight

section in urban and rural areas. In addition, the annual average daily traffic and type of accident played a key role in the safety of intersections in urban and rural areas, respectively. Finally, the type of accident and weekday had the most impact on safety in other sections of urban and rural areas.

For future research work, it is recommended to investigate the effectiveness of other artificial intelligence techniques to evaluate transportation safety, which is worth studying, and to extend the study to a five-year reporting period. In addition, the evaluation of the effects of geometric issues (inconsistencies) or pavement deterioration on safety analysis can be considered another important factor in rural environments in future works. Furthermore, it will be advisable to evaluate homogeneous road segments without reported crashes and to compare them with the used dataset to better understand the analyzed phenomenon by coupling artificial intelligence (AI) and machine learning (ML) techniques with other classic techniques, such as the naïve Bayesian (NB) algorithm and structural equation models (SEM).

Author Contributions: All authors have read and agreed to the published version of the manuscript. G.G., S.S.H. (Sina Shaffiee Haghsheenas) and S.S.H. (Sami Shaffiee Haghsheenas) were responsible for conceptualization and methodology. G.G. and A.V. analyzed the study context and extracted the dataset. A.V., V.A., and A.S.H. performed supervision, review and editing. S.S.H. (Sina Shaffiee Haghsheenas) and S.S.H. (Sami Shaffiee Haghsheenas) carried on the statistical analysis.

Funding: This research received no external funding.

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

References

1. Blas, E.; Kurup, A.S. (Eds.) *Equity, Social Determinants and Public Health Programmes*; World Health Organization: Geneva, Switzerland, 2010.
2. Li, Y.; Ma, D.; Zhu, M.; Zeng, Z.; Wang, Y. Identification of significant factors in fatal-injury highway crashes using genetic algorithm and neural network. *Accid. Anal. Prev.* **2018**, *111*, 354–363. [[CrossRef](#)] [[PubMed](#)]
3. Jovanis, P.P.; Chang, H.L. Modeling the relationship of accidents to miles traveled. *Transp. Res. Board* **1986**, *1068*, 42–51.
4. Poch, M.; Mannering, F. Negative binomial analysis of intersection-accident frequencies. *J. Transp. Eng.* **1996**, *122*, 105–113. [[CrossRef](#)]
5. Shankar, V.; Milton, J.; Mannering, F. Modeling accident frequencies as zero-altered probability processes: An empirical inquiry. *Accid. Anal. Prev.* **1997**, *29*, 829–837. [[CrossRef](#)]
6. Milton, J.; Mannering, F. The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. *Transportation* **1998**, *25*, 395–413. [[CrossRef](#)]
7. Hadi, M.A.; Aruldas, J.; Chow, L.; Wattleworth, J.A. Estimating safety effects of cross-section design for various highway types using negative binomial regression. *Transp. Res. Rec.* **1995**, *1500*, 169–177.
8. Miaou, S.P. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. *Accid. Anal. Prev.* **1994**, *26*, 471–481. [[CrossRef](#)]
9. Lee, J.; Mannering, F. Impact of roadside features on the frequency and severity of run-off-roadway accidents: An empirical analysis. *Accid. Anal. Prev.* **2020**, *34*, 149–161. [[CrossRef](#)]
10. Lord, D.; Mannering, F. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transp. Res. Part A Policy Pract.* **2010**, *44*, 291–305. [[CrossRef](#)]
11. Mannering, F.L.; Shankar, V.; Bhat, C.R. Unobserved heterogeneity and the statistical analysis of highway accident data. *Anal. Methods Accid. Res.* **2016**, *11*, 1–16. [[CrossRef](#)]
12. Mannering, F. Temporal instability and the analysis of highway accident data. *Anal. Methods Accid. Res.* **2018**, *17*, 1–13. [[CrossRef](#)]
13. Park, E.S.; Lord, D. Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity. *Transp. Res. Rec.* **2007**, *2019*, 1–6. [[CrossRef](#)]
14. El-Basyouny, K.; Sayed, T. Accident prediction models with random corridor parameters. *Accid. Anal. Prev.* **2009**, *41*, 1118–1123. [[CrossRef](#)] [[PubMed](#)]
15. Anastasopoulos, P.C.; Mannering, F.L. A note on modeling vehicle accident frequencies with random-parameters count models. *Accid. Anal. Prev.* **2009**, *41*, 153–159. [[CrossRef](#)] [[PubMed](#)]

16. Abduljabbar, R.; Dia, H.; Liyanage, S.; Bagloee, S.A. Applications of artificial intelligence in transport: An overview. *Sustainability* **2019**, *11*, 189. [[CrossRef](#)]
17. Karlaftis, M.G.; Golias, I. Effects of road geometry and traffic volumes on rural roadway accident rates. *Accid. Anal. Prev.* **2002**, *34*, 357–365. [[CrossRef](#)]
18. Abdel-Aty, M.; Keller, J. Exploring the overall and specific crash severity levels at signalized intersections. *Accid. Anal. Prev.* **2005**, *37*, 417–425. [[CrossRef](#)]
19. Park, Y.J.; Saccomanno, F.F. Collision frequency analysis using tree-based stratification. *Transp. Res. Rec.* **2005**, *1908*, 121–129. [[CrossRef](#)]
20. Yuan, F.; Cheu, R.L. Incident detection using support vector machines. *Transp. Res. Part C Emerg. Technol.* **2003**, *11*, 309–328. [[CrossRef](#)]
21. Yao, B.; Hu, P.; Zhang, M.; Jin, M. A support vector machine with the tabu search algorithm for freeway incident detection. *Int. J. Appl. Math. Comput. Sci.* **2014**, *24*, 397–404. [[CrossRef](#)]
22. Siddiqui, C.; Abdel-Aty, M.; Huang, H. Aggregate nonparametric safety analysis of traffic zones. *Accid. Anal. Prev.* **2012**, *45*, 317–325. [[CrossRef](#)] [[PubMed](#)]
23. Li, Z.; Liu, P.; Wang, W.; Xu, C. Using support vector machine models for crash injury severity analysis. *Accid. Anal. Prev.* **2012**, *45*, 478–486. [[CrossRef](#)]
24. Chen, C.; Zhang, G.; Qian, Z.; Tarefder, R.A.; Tian, Z. Investigating driver injury severity patterns in rollover crashes using support vector machine models. *Accid. Anal. Prev.* **2016**, *90*, 128–139. [[CrossRef](#)] [[PubMed](#)]
25. Holland, J.H. *Adaptation in Natural and Artificial Systems*; University of Michigan Press: Ann Arbor, MI, USA, 1975.
26. Arbis, D.; Dixit, V.V. Game theoretic model for lane changing: Incorporating conflict risks. *Accid. Anal. Prev.* **2019**, *125*, 158–164. [[CrossRef](#)] [[PubMed](#)]
27. Kim, Y.H.; Yoon, Y.; Geem, Z.W. A comparison study of harmony search and genetic algorithm for the max-cut problem. *Swarm Evol. Comput.* **2019**, *44*, 130–135. [[CrossRef](#)]
28. Aryafar, A.; Mikaeil, R.; Shafiee Haghshenas, S.; Shafiei Haghshenas, S. Utilization of soft computing for evaluating the performance of stone sawing machines, Iranian Quarries. *Int. J. Min. Geo-Eng.* **2018**, *52*, 31–36.
29. Fakharian, P.; Naderpour, H.; Haddad, A.; Rafiean, A.H.; Eidgahee, D.R. A proposed model for compressive strength prediction of FRP-confined rectangular column in terms of Genetic expression Programming (GEP). *Concr. Res.* **2018**, *11*, 5–18.
30. Haghshenas, S.S.; Faradonbeh, R.S.; Mikaeil, R.; Haghshenas, S.S.; Taheri, A.; Saghatforoush, A.; Dormishi, A. A new conventional criterion for the performance evaluation of gang saw machines. *Measurement* **2019**, *146*, 159–170. [[CrossRef](#)]
31. Pinto, B.Q.; Ribeiro, C.C.; Rosseti, I.; Noronha, T.F. A biased random-key genetic algorithm for routing and wavelength assignment under a sliding scheduled traffic model. *J. Glob. Optim.* **2020**, *77*, 949–973. [[CrossRef](#)]
32. Luan, J.; Yao, Z.; Zhao, F.; Song, X. A novel method to solve supplier selection problem: Hybrid algorithm of genetic algorithm and ant colony optimization. *Math. Comput. Simul.* **2019**, *156*, 294–309. [[CrossRef](#)]
33. Hosseini, S.M.; Ataei, M.; Khalokakaie, R.; Mikaeil, R.; Haghshenas, S.S. Study of the effect of the cooling and lubricant fluid on the cutting performance of dimension stone through artificial intelligence models. *Eng. Sci. Technol. Int. J.* **2020**, *23*, 71–81. [[CrossRef](#)]
34. Salemi, A.; Mikaeil, R.; Haghshenas, S.S. 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.* **2017**, *22*, 1978–1990. [[CrossRef](#)]
35. Slowik, A.; Kwasnicka, H. Nature inspired methods and their industry applications—Swarm intelligence algorithms. *IEEE Trans. Ind. Inform.* **2018**, *14*, 1004–1015. [[CrossRef](#)]
36. Dulebenets, M.A. A Comprehensive Evaluation of Weak and Strong Mutation Mechanisms in Evolutionary Algorithms for Truck Scheduling at Cross-Docking Terminals. *IEEE Access* **2018**, *6*, 65635–65650. [[CrossRef](#)]
37. Brezočnik, L.; Fister, I.; Podgorelec, V. Swarm intelligence algorithms for feature selection: A review. *Appl. Sci.* **2018**, *8*, 1521. [[CrossRef](#)]
38. Anandakumar, H.; Umamaheswari, K. A bio-inspired swarm intelligence technique for social aware cognitive radio handovers. *Comput. Electr. Eng.* **2018**, *71*, 925–937. [[CrossRef](#)]
39. Zhao, X.; Wang, C.; Su, J.; Wang, J. Research and application based on the swarm intelligence algorithm and artificial intelligence for wind farm decision system. *Renew. Energy* **2019**, *134*, 681–697. [[CrossRef](#)]

40. Dulebenets, M.A. An Adaptive Island Evolutionary Algorithm for the berth scheduling problem. *Memetic Comput.* **2020**, *12*, 51–72. [\[CrossRef\]](#)
41. Kandiri, A.; Golafshani, E.M.; Behnood, A. Estimation of the compressive strength of concretes containing ground granulated blast furnace slag using hybridized multi-objective ANN and salp swarm algorithm. *Constr. Build. Mater.* **2020**, *248*, 118676. [\[CrossRef\]](#)
42. Mikaeil, R.; Haghshenas, S.S.; Haghshenas, S.S.; Ataei, M. Performance prediction of circular saw machine using imperialist competitive algorithm and fuzzy clustering technique. *Neural Comput. Appl.* **2018**, *29*, 283–292. [\[CrossRef\]](#)
43. Mikaeil, R.; Haghshenas, S.S.; Hoseinie, S.H. Rock penetrability classification using artificial bee colony (ABC) algorithm and self-organizing map. *Geotech. Geol. Eng.* **2018**, *36*, 1309–1318. [\[CrossRef\]](#)
44. 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*, 3779–3791. [\[CrossRef\]](#)
45. Mikaeil, R.; Haghshenas, 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*, 1099–1113. [\[CrossRef\]](#)
46. Mikaeil, R.; Beigmohammadi, M.; Bakhtavar, E.; Haghshenas, S.S. Assessment of risks of tunneling project in Iran using artificial bee colony algorithm. *SN Appl. Sci.* **2019**, *1*, 1711. [\[CrossRef\]](#)
47. Dormishi, A.; Ataei, M.; Mikaeil, R.; Khalokakaei, R.; Haghshenas, S.S. Evaluation of gang saws' performance in the carbonate rock cutting process using feasibility of intelligent approaches. *Eng. Sci. Technol. Int. J.* **2019**, *22*, 990–1000. [\[CrossRef\]](#)
48. 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*, 8545–8559.
49. Fiorini Morosini, A.; Shaffiee Haghshenas, S.; Shaffiee Haghshenas, 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*, 3029. [\[CrossRef\]](#)
50. Sarkar, S.; Vinay, S.; Raj, R.; Maiti, J.; Mitra, P. Application of optimized machine learning techniques for prediction of occupational accidents. *Comput. Oper. Res.* **2019**, *106*, 210–224. [\[CrossRef\]](#)
51. Liu, Y.; Zou, B.; Ni, A.; Gao, L.; Zhang, C. Calibrating microscopic traffic simulators using machine learning and particle swarm optimization. *Transp. Lett.* **2020**, *1*–13. [\[CrossRef\]](#)
52. Kennedy, J.; Eberhart, R. Particle swarm optimization. In Proceedings of the ICNN'95-International Conference on Neural Networks, Perth, WA, Australia, 27 November–1 December 1995; pp. 1942–1948.
53. Poli, R.; Kennedy, J.; Blackwell, T. Particle swarm optimization. *Swarm Intell.* **2007**, *1*, 33–57. [\[CrossRef\]](#)
54. Frank, L.R.; Ferreira, Y.M.; Julio, E.P.; Ferreira, F.H.C.; Dembogurski, B.J.; Silva, E.F. Multilayer Perceptron and Particle Swarm Optimization Applied to Traffic Flow Prediction on Smart Cities. In *International Conference on Computational Science and Its Applications*; Springer: Cham, Switzerland, 2019; pp. 35–47.
55. Chen, L.; Monteiro, T.; Wang, T.; Marcon, E. Design of shared unit-dose drug distribution network using multi-level particle swarm optimization. *Health Care Manag. Sci.* **2019**, *22*, 304–317. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Liu, P.; Xie, M.; Bian, J.; Li, H.; Song, L. A Hybrid PSO–SVM Model Based on Safety Risk Prediction for the Design Process in Metro Station Construction. *Int. J. Environ. Res. Public Health* **2020**, *17*, 1714. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Tharwat, A.; Elhoseny, M.; Hassanien, A.E.; Gabel, T.; Kumar, A. Intelligent Bézier curve-based path planning model using Chaotic Particle Swarm Optimization algorithm. *Clust. Comput.* **2019**, *22*, 4745–4766. [\[CrossRef\]](#)
58. 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*, 3125–3143. [\[CrossRef\]](#)
59. Lloyd, S. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **1982**, *28*, 129–137. [\[CrossRef\]](#)
60. Feng, X.; Li, S.; Yuan, C.; Zeng, P.; Sun, Y. Prediction of slope stability using naive Bayes classifier. *KSCE J. Civ.* **2018**, *22*, 941–950. [\[CrossRef\]](#)
61. Hosseini, S.M.; Ataei, M.; Khalokakaei, R.; Mikaeil, R.; Haghshenas, S.S. Investigating the Role of the Cooling and Lubricant Fluids on the Performance of Cutting Disks (Case Study: Hard Rocks). *Rud. Geološko Naft. Zb.* **2019**, *34*.

62. Pirouz, B.; Shaffiee Haghshenas, S.; Shaffiee Haghshenas, S.; Piro, P. Investigating a serious challenge in the sustainable development process: Analysis of confirmed cases of COVID-19 (new type of coronavirus) through a binary classification using artificial intelligence and regression analysis. *Sustainability* **2020**, *12*, 2427. [[CrossRef](#)]
63. Mussone, L.; Bassani, M.; Masci, P. Analysis of factors affecting the severity of crashes in urban road intersections. *Accid. Anal. Prev.* **2019**, *103*, 112–122. [[CrossRef](#)]
64. Dutta, N.; Fontaine, M.D. Improving freeway segment crash prediction models by including disaggregate speed data from different sources. *Accid. Anal. Prev.* **2019**, *132*, 1–16. [[CrossRef](#)]
65. 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. *J. Min. Environ* **2020**, *10*, 543–557.
66. Shaffiee Haghshenas, S.; Pirouz, B.; Shaffiee Haghshenas, S.; Pirouz, B.; Piro, P.; Na, K.S.; Geem, Z.W. Prioritizing and Analyzing the Role of Climate and Urban Parameters in the Confirmed Cases of COVID-19 Based on Artificial Intelligence Applications. *Int. J. Env. Res. Public Health* **2020**, *17*, 3730. [[CrossRef](#)] [[PubMed](#)]
67. Mikaeil, R.; Bakhshinezhad, H.; Haghshenas, S.S.; Ataei, M. Stability analysis of tunnel support systems using numerical and intelligent simulations (case study: Kouhin Tunnel of Qazvin-Rasht Railway). *Rud. Geološko Naft. Zb.* **2019**, *34*, 1–10. [[CrossRef](#)]
68. Mirjalili, S. Genetic algorithm. In *Evolutionary Algorithms and Neural Networks*; Springer: Cham, Switzerland, 2019; pp. 43–55.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).