IJTE, Vol. 11, No. 2, Autumn 2023, pp. 1471-1489
Research Paper

A Genetic Algorithm Based Reliability Assessment of Crash Modification Factors Calculated by Full Bayesian Method

Farzad Akbarinia[1,*], Hamidreza Behnood[2]

Received: 2023/05/28 *Accepted: 2023/08/20*

Abstract

Given the importance of reducing crashes, as well as the implementation of principled and practical road safety measures, it is necessary to calculate the Crash Modification Factors (CMFs), which is the main factor in road safety effectiveness analysis, with high accuracy. Therefore, reliability assessment is of great importance and necessity for calculating CMFs. In this study, a method for evaluating the reliability of CMFs using metaheuristic Genetic Algorithm (GA) is presented. The proposed model is defined based on the before-after study method with the comparison group and based on the Full Bayesian (FB) method for calculating CMFs. The Monte Carlo Markov Chain (MCMC) has been applied for calculating posterior distributions as a sampling method that allows the simulation of posterior samples from complex distributions. Crash data, categorized into total crashes and fatal-injury crashes, were collected from the city of Karaj and had a period of 5 years (2016-2020). The remedial action of signalizing the intersections along with the installation of the counters was considered as the treatment considered in the study. Results show that the CMF value for remedial action of signalizing intersections with counters, did not have a significant impact on reducing total crashes (CMF=1.07). On the other hand, by evaluating the CMF values calculated for fatal-injury crashes, it is determined that the calculated CMF is approximately equal to 0.75, which indicates the positive effect of the remedial action and reduction of fatal-injury crashes. In addition, according to the proposed GA-based rating system, CMF values for fatal-injury crashes have the highest rank, which indicates a very high reliability for the calculated CMF values. Therefore, it is possible to confidently take the remedial action of signalizing intersections with the installation of a counter as an effective measure to reduce the number of fatal-injury crashes.

Keywords: Crash Modification Factor, Genetic Algorithm, Full Bayesian, Monte Carlo Markov Chain

[*] Corresponding author. E-mail: farzadakbarinia@gmail.com
[1] Department of Transportation Planning, Faculty of Engineering, Imam Khomeini International University, Qazvin, Iran
[2] Department of Transportation Planning, Faculty of Engineering, Imam Khomeini International University, Qazvin, Iran

1. Introduction

One of the main concerns of road safety practitioners has been the design and implementation of road safety measures in various areas with the aim of reducing the number of crashes. The most important criterion introduced in recent years to evaluate the success and effectiveness of remedial measures in reducing crashes has been the Crash Modification Factor (CMF) (National Research Council (US), 2010). This factor indicates the amount of decrease or increase in crashes after the implementation of remedial measures. Mainly for calculating, evaluating and developing CMFs, various data of crashes, road geometric features, traffic characteristics, etc. have been needed, and in this regard, the use of numerical models is considered by researchers.

Due to various factors affecting CMF calculations in different methods, CMF computational values have been uncertain and unreliable. This issue, based on the type and severity of crashes as well as the unique conditions of each improvement site, has caused more uncertainty and has made the CMF values computationally optimistic and far from reality. Given the importance of reducing crashes, as well as the implementation of principled and practical remedial measures, it is necessary to calculate the CMF, which is the main factor in road safety effectiveness analysis, with high accuracy. Therefore, reliability assessment is of great importance and necessity for calculating CMFs.

Due to the importance of CMF reliability assessment, in this study, a method for evaluating the reliability of CMFs using metaheuristic genetic algorithm is presented. The proposed model is defined based on the before-after study method with the comparison group and based on the Full Bayesian (FB) method for calculating CMFs. In this study, a ranking for the reliability of CMFs based on crash severities is presented. Due to the superior features of statistical modeling techniques and its computational capacity, some researchers have modeled and conducted CMF studies using the FB method (e.g. Li et al., 2008; Lan et al., 2009; Persaud et al., 2010; El-Basyouny and Sayed, 2009-2012; Ahmed et al., 2015; Pu et al., 2020; and Ali et al., 2021). Based on researches conducted by various researchers (Brimley et al., 2012; Lubliner and Schrock, 2012; Abdel-Aty et al., 2014; Park et al., 2015; and Wang et al., 2015), CMF calculation based on before-after methods may be unreliable and according to the different conditions of treatment and control sites, computational values for CMFs can be subject to bias. In this regard, various studies have been conducted in which the sensitivity of different parameters of each site in determining CMFs has been investigated, such as time conditions (Park et al., 2015; Sacchi et al., 2014; Wang et al., 2015a), Traffic volume (Sacchi and Sayed, 2014; Wang and Abdel-Aty, 2014), area type (Wang and Abdel-Aty, 2014), and speed limit (Lee et al., 2015).

Relatively limited research has been conducted on the subject of evaluating the reliability and application of meta-heuristic algorithms, especially genetic algorithms. Cheng and Li (2008) evaluated the reliability of structures using genetic algorithms. In these studies, a new class of artificial neural network-based genetic algorithms (ANN-GA) for reliability analysis of structures was developed. Tun et al. (2016) investigated and evaluated the reliability of embankments and slopes using genetic algorithm. In this research, a genetic algorithm has been developed to solve the optimization problem by considering the limit equilibrium method for searching multiple critical failures. The results of this study showed that the proposed model based on the use of genetic algorithm has a high ability to assess the reliability of slope stability. Munyazikwiye et al. (2017) developed a mathematical model for vehicle frontal crashes. To estimate and optimize the model parameters, a genetic algorithm approach was proposed. The results of this study showed that the developed model can accurately reproduce the real kinematic results from the crash test.

In Amiri et al. (2020), Run-off-road (ROR) crashes focusing on two factors of drivers' age (over 65 years) and collisions with hard objects were analyzed using two methods of genetic algorithm and artificial neural network. Two types of artificial intelligence (AI) techniques, including hybrid intelligent genetic algorithm and artificial neural network (ANN), were used for the studies. Although the results show that the developed ANN performs better than the hybrid intelligent genetic algorithm, the hybrid approach is more capable of predicting high-severity crashes. Ospina-Mateus et al. (2021) studied and analyzed motorcycle crashes on Bogota roads in Colombia using the genetic algorithm. In this study, both genetic and simulated annealing algorithms were applied in relation to the extraction rules (support, reliability, enhancement and comprehensibility) in accordance with the objectives of the problem. The results showed that the use of genetic algorithm and also the implementation of the model based on the hybrid algorithm has improved the results by 21 percent compared to other methods. Mrowicki et al. (2021) evaluated the effect of speed in crashes using the genetic algorithm. In this study, a new approach to pre-crash velocity development using the genetic algorithm driven multi-agent (GAMA) was developed. Howlader et al. (2023) have evaluated the impact of the protected right turn phase in signalized intersection crashes. In this research, the effectiveness of protected right turn signal phasing has been done using the method of heterogeneous count data models with experimental and incomplete techniques. Empirical Bayes method is based on Poisson-Gamma models and full Bayes is based on Poisson-Lognormal intervention models. The results showed the reduction of accidents in the protected right turn mode by about 87 and 91 percent for normal intersections and T-shaped intersections, respectively. Al-Marafi and Somasundaraswaran (2023) have evaluated the results of CMF calculation with observational and cross-sectional methods. The results show that among the BA methods, two Full Bayes and Empirical Bayes methods have higher accuracy, at the same time, determining the exact place and time of the implementation of the correction, performing more than one corrective action is one of the limitations of these methods. Bin Tahir et al. (2023) have presented a framework for evaluating CMF calculation methods by hypothetical treatments with known ground truth and actual real-world treatments. Empirical Bayes, Simulated Empirical Bayes, and Full Bayes have been investigated. The results show that all methods can identify the ground truth of hypothetical treatments, but the full Bayes method provides better results. Vinayaraj and Perumal (2023) have calculated CMF and SPF in 21 fields with 78 approaches. The obtained results show the influence of factors such as the diameter of the field island, average daily traffic, entry and exit times in the occurrence of accidents. According to the type of data, the reliability of these results can be one of the limitations of the research.

Before-after study methods introduce two basic problems into CMF calculations in the usual method, which is the limitation in the number of samples and as a result the small number of data that causes errors, and the second is the non-random selection of sites. The calculated CMF values based on each of the parameters of the type and severity of the accidents, as well as the characteristics of the modified locations, have uncertainty. In this research, a model based on before-after methods has been implemented using the full Bayesian function FB, which has special features compared to other before-after methods such as empirical Bayes EB. The genetic algorithm method has been used to develop the model in order to evaluate the reliability. As one of the results and outputs of this research, the scoring and ranking of CMF reliability was based on the type and severity of accidents, which is innovative based on the proposed method and model.

2. Methods

2.1. Full Bayesian Method

In this research, the Full Bayes model has been used to conduct studies and calculate CMFs. The main idea in Bayesian models is to use probability distributions for various parameters affecting the model performance. In the following, the main theory of the FB model is presented using comparison groups to achieve the CMFs (Avelar et al., 2021):

- The rate of change in the number of annual crashes in urban areas has different probability distributions, based on which CMF analytical methods are presented. One of the applied and up-to-date study models is the use of Poisson-gamma mixed distribution model.
- Poisson-gamma mixed distribution from a technical and computational point of view has resulted in negative binomial distribution that in different models can be used instead of Poisson-gamma mixed distribution.
- In this study, the CMF evaluation model was a model based on the Poisson-gamma mixed distribution for the FB before-after and using comparison groups.
- Before presenting the theoretical foundations, the main parameters in the model are introduced as follows:
  - y_{it}: Number of crashes on site *i* in year *t*
  - *k*: Number of covariates
  - $\mathbf{X}_{it}$: Vector of covariates with *K*+1 dimension; $\mathbf{X}_{it} = (1, X_{1it} \cdots X_{Kit})$
  - $\boldsymbol{\beta}$: Regression coefficients with *K*+1 dimension; $\boldsymbol{\beta} = (\beta_0, \beta_1 \cdots \beta_K)'$
  - $\boldsymbol{\upsilon}_{it}$: Annual random effects vector at site *i* and year *t*

Based on the introduced parameters, the model in question has the distribution of annual crashes as follows:

$$y_{it}|\upsilon_{it}, \beta \sim \text{Poisson}(\mu_{it}) \quad (1)$$

According to Equation 1, the distribution of the number of annual crashes y_{it} has a mean of μ_{it} with the Poisson distribution (Equation 2). Also, due to the mixed model, the parameter υ_{it} has a gamma distribution as described in Equation 3.

$$\mu_{it} = \nu_{it} \exp(X_{it}.\beta) \quad (2)$$

$$\nu_{it} \sim \text{Gamma}(\eta, {}^{1}/_{\eta}) \quad (3)$$

According to the explanations provided, the y_{it} has a negative binomial marginal distribution that has a mean and variance as described in 4:

$$\begin{cases} \text{mean } \lambda_i \\ \text{Variance } \lambda_i[1+\lambda_i|\eta] \end{cases} \rightarrow \lambda_i = \exp(X_i\beta) \quad (4)$$

According to the model, the $\mathbf{X}_{it}$ vector covariates are presented in Equation 9. Based on the description and covariates of Equations 1 to 4, the relation of the mean μ_{it} crashes can be rewritten as described in 5:

$$\begin{cases} X_{1it} = Trt_i \\ X_{2it} = Time \\ X_{3it} = Trt_i \times Time \\ X_{4it} = I[t > t_{0i}] \\ X_{5it} = Trt_i \times I[t > t_{0i}] \\ \quad X_{6it}, \dots, X_{Kit}\text{: Intersection characteristic variables} \end{cases} \quad (5)$$

$$\begin{aligned} \mu_{it} = \nu_{it}\exp(\beta_0 + \beta_1 Trt_i + \beta_2 time + \beta_3 Trt_i \times time \\ + \beta_4 I[t > t_{0i}] + \beta_5 Trt_i \times I[t > t_{0i}] \\ + \beta_6 X_{6it} + \cdots + \beta_K X_{Kit}) \end{aligned} \quad (6)$$

Given the different values of the **X** vector, which have different values based on the base year of the remedial actions and the period before and after the remedial action, according to Equation 5, Equation 6 can be used to calculate μ_{it} based on the before or after period as well as the type of treatment or comparison site as formulated in 7.

$$\begin{cases} (\mu_{it})_{comp,B} = \nu_{it}\exp(\beta_0 + \beta_2 time + \beta_6 X_{6it} + \cdots + \beta_K X_{Kit}) \\ (\mu_{it})_{comp,A} = \nu_{it}\exp(\beta_0 + \beta_4 + \beta_2 time + \beta_6 X_{6it} + \cdots + \beta_K X_{Kit}) \\ (\mu_{it})_{Trt,B} = \nu_{it}\exp(\beta_0 + \beta_1 + (\beta_2 + \beta_3)time + \beta_6 X_{6it} + \cdots + \beta_K X_{Kit}) \\ (\mu_{it})_{Trt,A} = \nu_{it}\exp(\beta_0 + \beta_1 + \beta_4 + \beta_5 + (\beta_2 + \beta_3)time + \beta_6 X_{6it} + \cdots + \beta_K X_{Kit}) \end{cases} \quad (7)$$

This model evaluates and calculates the CMF using the before-after method based on the FB method as well as treatment and comparison sites. The proposed model is developed based on different possible distributions for different parameters. In this model, in order to find the parameters of the considered distributions as well as the values of the **β** vector, it is necessary to obtain the unknown parameters based on the prior distributions and the behavior of the model based on the available data. For this purpose, it is necessary to use simulation-based methods that obtain the values of unknown parameters by successive sampling. In this study, the Monte Carlo Markov Chain (MCMC) has been applied. The MCMC method is one of the common techniques for calculating posterior distributions. MCMC is a sampling method that allows the simulation of posterior samples from complex distributions.

The MCMC sampling method uses a Markov chain model in a random method, so that the distribution of the next sample depends only on the current sample. As the algorithm evaluates more samples, the samples approximate the closer probability distribution (PDF) function. Specifically, starting with a custom prototype (current sample), a new candidate sample is prepared from a proposal.

2.2. Calculation Steps

In the following, the process of performing calculations and the main steps to create and implement the model are presented. The steps of performing calculations and executing the model include the 8 main steps as follows:

Step 1: In the first step, the data set that includes comprehensive information of the study area (number of annual crashes per site, geometric and traffic characteristics of each site including AADT, speed limit, length of route, etc.) are collected for periods before and after the remedial action.

Step 2: Categorize the initial data based on the type of site and the study period; At this stage, after obtaining the initial data, according to the type of site, which is divided into two categories of treatment sites and comparison sites, the data related to the period before the remedial action and after it are classified (T: Treatment site, C: Comparison site, A: Time period after remedial action, B: Time period before remedial action).

Step 3: In this step, the model is formed based on the proposed theory. In this step, based on the relationships provided for each of the four main conditions (Equation 7), the parameters of the considered distributions and the values of the **β** vector are calculated using the MCMC method and the amount of computational error is calculated based on MSE and RMSE tests. It should be noted that in the MCMC method, considering 25,000 sampling times and removing the first 5000 samples, the best values for the unknown parameters are calculated.

Step 4: After obtaining the results of the previous step, the mean values of crash frequencies for each site and for each period are calculated, based on the estimated values of μ_{CB}, μ_{CA}, μ_{TB}, and μ_{TA}.

Step 5: Using Equation 8, the prior distribution of the predicted crash frequency ratios before and after the period is calculated for the comparison group (comparison ratio).

$$\mathrm{R}_{\mathrm{C(g)}} = \frac{\mu_{\mathrm{CA(g)}}}{\mu_{\mathrm{CB(g)}}} \tag{8}$$

Step 6: The prior distribution of crashes is predicted, which is calculated for the group of treatment sites based on Equation 9 without considering the remedial action in the period after the treatment.

$$\pi_{\mathrm{(g)}} = \mu_{\mathrm{TB(g)}}\mathrm{R}_{\mathrm{C(g)}} \tag{9}$$

Step 7: Calculate CMF values; at this stage, the values of the Crash Modification Factors are calculated based on Equation 10.

$$\theta = \frac{\sum_{\mathrm{g=1}}^{\mathrm{G}} \mu_{\mathrm{TA(g)}}}{\sum_{\mathrm{g=1}}^{\mathrm{G}} \pi_{\mathrm{(g)}}} = \frac{\sum_{\mathrm{g=1}}^{\mathrm{G}} \mu_{\mathrm{TA(g)}}}{\sum_{\mathrm{g=1}}^{\mathrm{G}} \{\mu_{\mathrm{TB(g)}}\mathrm{R}_{\mathrm{C(g)}}\}} \tag{10}$$

Step 8: Assess reliability; at this stage, a genetic algorithm is used to evaluate the reliability. Therefore, considering different conditions for the main parameters involved in the $\mathbf{X}_{it}$ variable vector, we consider four main conditions for forming the μ_{it} model equations. According to the provided explanations, the values of **X** vector (X1, X2… X5) have different values depending on the type of site and the period in question. Also, the values of X_{it6}… X_{itk} represent the traffic and geometric characteristics of the sites in question, based on this, we consider 4 different situations as follows to calculate the CMF:

- Calculation based on the main parameters of the site type and study period: $\mathbf{X} = (X1, X2, X3, X4, X5)$
- Calculate by applying the AADT parameter: $\mathbf{X} = (X1, X2, X3, X4, X5, X6)$
- Calculate by applying the maximum limited speed (LMS): $\mathbf{X} = (X1, X2, X3, X4, X5, X6, X7)$
- Calculate by applying the rout length: $\mathbf{X} = (X1, X2, X3, X4, X5, X6, X7, X8)$

CMF values are recalculated to evaluate the reliability using genetic algorithm. In the reliability evaluation model, the genetic metaheuristic algorithm is used to evaluate the maximum and minimum values of CMF based on different sampling modes (considering different model parameters for calculating CMF). Therefore, to evaluate the reliability using GA, first the objective function must be determined, which according to the description provided includes the CMF calculation function.

Once the objective function is specified, the model is executed once to achieve the maximum CMF values and once to achieve the minimum values. It should be noted that the GA algorithm method, based on the creation of the initial community and the operators of selection, multiplication and mutation, has the ability to search in different spaces to achieve the optimal values of the objective function.

Based on the evaluation and comparison between the calculated CMF values of the model with the maximum and minimum values obtained from the GA algorithm, the reliability of computational CMFs is discussed. Based on the proposed theory and structure, the model algorithm was created and coded using a suitable programming platform. The flowchart of Figure 1 presents the main algorithm and study design.

2.3. Case Study

In this study, the performance of the crash prediction model and the reliability assessment of the CMF using the GA algorithm have been implemented in a case study. The study case in this study was the intersections of Karaj city in Iran. Crash data collected from the city of Karaj had a period of 5 years (beginning: 2016, end: 2020). The remedial action of signalizing the intersections along with the installation of the counters was considered as the treatment considered in the study. Treatment sites and comparison sites included 18 and 13 intersections respectively.

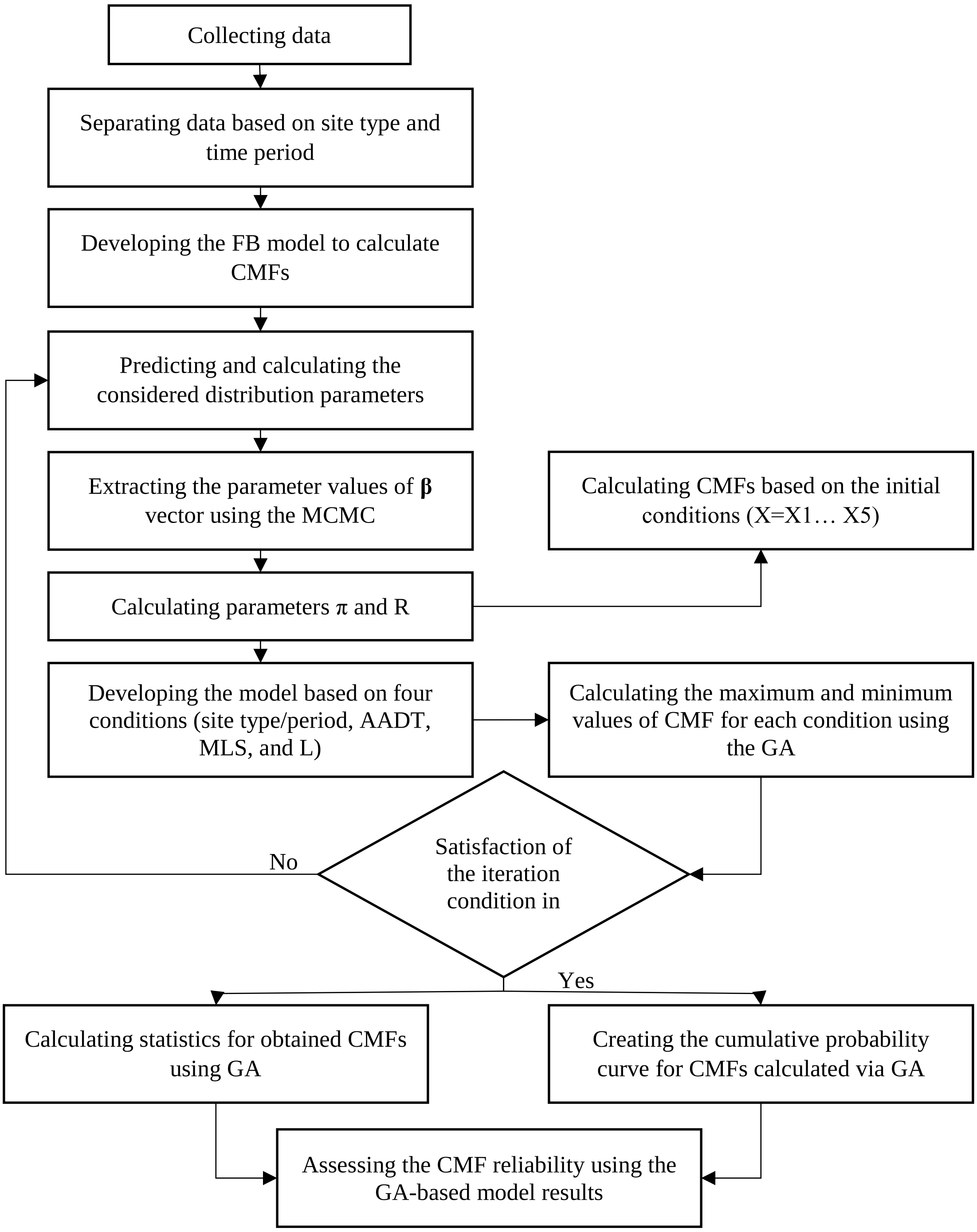

Figure 1. Study design

In order to initially evaluate and have a general understanding of the trend of crash changes based on the type of sites (treatment or comparison) and different years (periods before and after remedial action), total and fatal-injury crashes in all sites are shown in Table 1.

Table 1. Total and fatal-injury crashes in all sites

Site type	Before period		After period	
	Total crashes	Fatal-injury crashes	Total crashes	Fatal-injury crashes
treatment	3342	714	3813	304
comparison	2787	597	2643	392

3. Results

Based on the created model and applying the initial data from the study area (crash data of treatment and comparison sites), the values of CMF are calculated based on crash severities in the two sections of total crashes and fatal-injury crashes. The results of the model implementation include the three main parts of calculating the CMFs as follows:

- CMF calculation based on the proposed model and without implementing genetic algorithm (CMF calculation directly)
- CMF calculation based on GA algorithm
- Evaluation of CMF reliability and ranking them based on GA algorithm

3.1. CMF Calculation Directly

In this section, the results of model implementation for different modeling conditions and based on crash severities in two categories of total crashes and fatal-injury crashes are presented. According to the explanations provided about the relationships and the proposed model, in this model the values of the CMF for each level of crash severity are calculated under four different conditions (site type/period, AADT, MLS, and L). Table 2 presents the CMF values calculated under each of the four conditions for total crashes and fatal-injury crashes. In addition, the results of the FB model for calculating unknown coefficients at treatment and comparison sites (based on the main parameters of site type and study period) and the MCMC results for comparison sites are displayed in Table 3 and Figure 2, respectively.

Table 2. CMFs calculated based on different conditions for total and fatal-injury crashes

Condition	Calculated CMF	
	Total crashes	Fatal-injury crashes
X = (X1, X2, X3, X4, X5)	1.0576	0.75911
X = (X1, X2, X3, X4, X5, X6)	1.0774	0.75902
X = (X1, X2, X3, X4, X5, X6, X7)	1.0753	0.75058
X = (X1, X2, X3, X4, X5, X6, X7, X8)	1.0736	0.74517
Average CMF	1.07	0.75

Table 3. FB model results based on site type and study period

Before period for comparison sites					
	Mean	Std	CI95		Positive
Intercept	1.1477	0.69024	-0.057924	2.4439	0.94496
Beta(1)	1.1836	0.68661	-0.047966	2.4662	0.95296
Beta(2)	0.78603	0.15776	0.48011	1.0989	0.99996
Sigma2	0.054467	0.014769	0.03267	0.090377	1
After period for comparison sites					
	Mean	Std	CI95		Positive
Intercept	1.1625	0.68949	-0.057336	2.4665	0.95056
Beta(1)	1.1716	0.68912	-0.05383	2.4705	0.948
Beta(2)	0.78478	0.15742	0.47737	1.0976	1
Sigma2	0.054301	0.01455	0.032858	0.089246	1

Before period for treatment sites					
	Mean	**Std**	**CI95**		**Positive**
Intercept	1.0908	0.29366	0.51383	1.674	0.99996
Beta(1)	1.0867	0.29629	0.49683	1.6633	0.999
Beta(2)	1.0839	0.30047	0.4967	1.6757	0.99876
Beta(3)	0.027211	0.10973	-0.20017	0.28973	0.6458
Beta(4)	0.028193	0.10949	-0.19197	0.29267	0.6508
Sigma2	0.01236	0.0018631	0.0092425	0.016512	1
After period for treatment sites					
	Mean	**Std**	**CI95**		**Positive**
Intercept	0.55688	0.80125	-0.60324	2.4083	0.74572
Beta(1)	0.55703	0.79776	-0.57877	2.3978	0.74412
Beta(2)	0.55742	0.79722	-0.59894	2.3817	0.74624
Beta(3)	0.095498	0.38113	0.63032	1.0359	0.60548
Beta(4)	0.088225	0.38383	0.66206	1.0303	0.59668
Beta(5)	0.56616	0.80576	-0.564669	2.4365	0.74192
Beta(6)	0.55035	0.79968	-0.56017	2.4044	0.73772
Sigma2	0.097064	0.040557	0.040557	0.22342	1

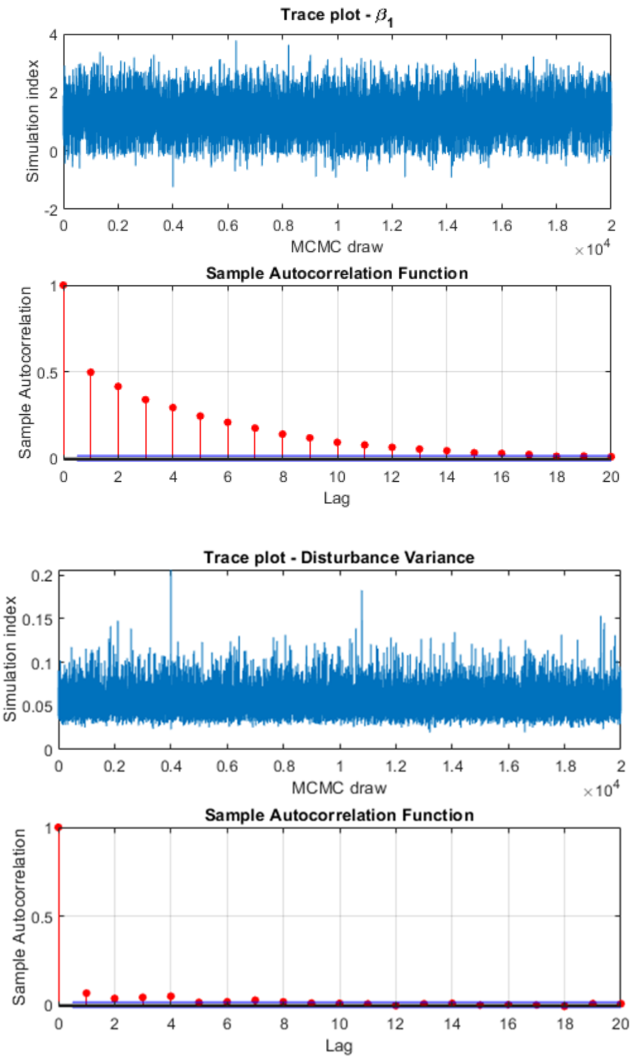

Figure 2. MCMC results for comparison sites: (a) Before period

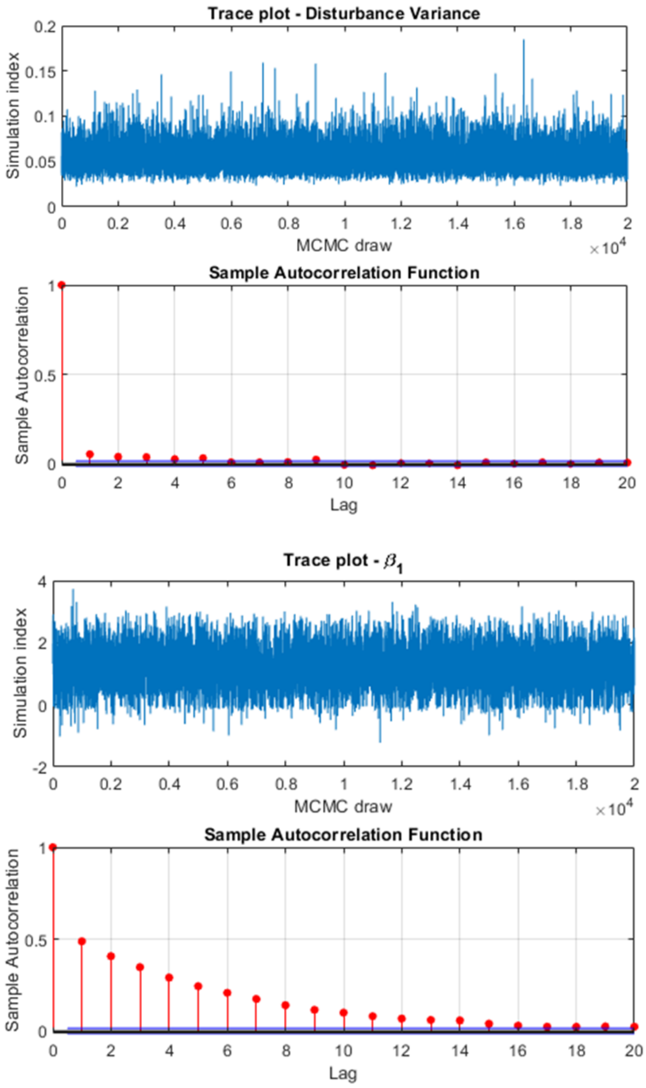

Figure 2. MCMC results for comparison sites: (b) After period

3.2. GA-based Reliability Assessment

In order to implement the model based on the GA algorithm, once the objective function is specified, the model is executed once to achieve the maximum CMF values and once to achieve the minimum values. It should be noted that the GA algorithm method based on the creation of the initial population and the operators of selection, multiplication and mutation, has the ability to search in different spaces to achieve optimal values of the objective function. In the proposed method for calculating the CMF reliability, the maximum and minimum values of CMF are calculated based on the GA algorithm and based on the evaluation and comparison between the calculated CMF values of the model with the maximum and minimum values obtained from the GA algorithm, the reliability rate for calculated CMFs are discussed. In order to implement the genetic algorithm, various initial parameters must be determined, which will be examined below. The first effective parameter in using this method is to determine the initial values or the initial sampling population for the model, which in the research model, the average values of CMF obtained from the model are presented as initial values. The second effective parameter was to determine the convergence range of the model, for which two control tools are used. The first lever is the confidence interval, which is set as a limit for stopping calculations. For example, the difference in the calculated value for the minimum or maximum CMF in two consecutive time steps or the difference in the initial value of the values calculated in each step can be an efficient measure for this purpose. The second lever was the number of iterations, which indicated the number of repetitions intended to achieve the goal (minimum or maximum CMF value). The third category parameters included initial values for population size, crossover rate and mutation rate. Population size, which determines the total number of possible states for production and calculation of CMFs, as well as crossover rate and mutation rate parameters based on the theory presented in section 2, indicate the percentage of impact and use of each of these operators on next generation population production. All the basic parameters required to set up a reliability assessment model based on the GA method are presented in Table 4.

Table 4. GA parameters in calculation of CMF reliability

Parameter	Value
Confidence interval	From the initial calculated CMF value equal to 0.6
Iteration	1000
Population size	1000
Crossover rate	0.8
Mutation rate	0.2

Based on the explanations provided, and taking into account the values of the parameters in Table 6, the models are implemented and the reliabilities of computational CMFs for total crashes and fatal-injury crashes are extracted, the results of which are presented below.

3.3. Reliability Assessment for Total Crashes

In this section, the results of the implementation of the GA model for total crashes are presented. According to the cases mentioned in the previous section and taking into account the parameters presented in Table 3, the model has been implemented in two modes as follows: (a) Objective function: Extraction of maximum CMF values, and (b) Objective function: Extraction of minimum CMF values. The diagrams in Figures 3 and 4 show the number of iterations of the GA computational model based on the objective functions (maximum and minimum CMF values) obtained from the implementation of the model under the two conditions.

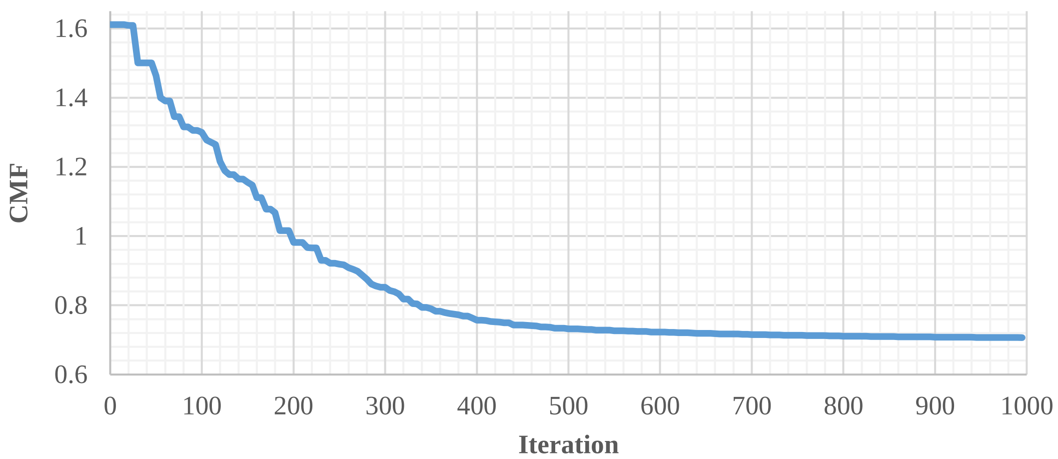

Figure 3. Calculation of minimum CMF values based on GA algorithm for total crashes

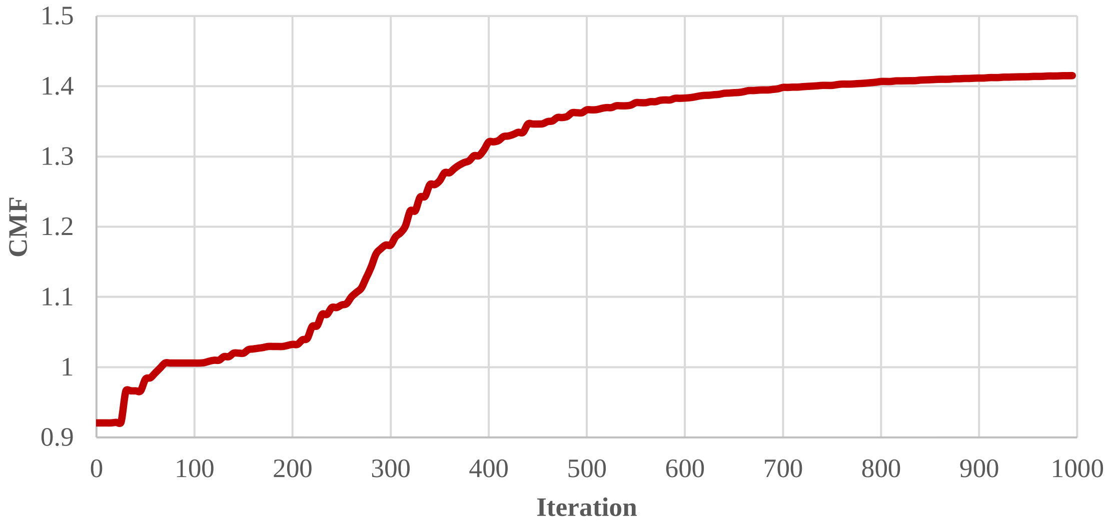

Figure 4. Calculation of maximum CMF values based on GA algorithm for total crashes

In order to evaluate the results more accurately, we removed the initial 10 percent of the CMF computational results (with the aim of initializing and generating the population to start solving the algorithm) and evaluated the statistical parameters obtained from the calculated values. Table 5 presents the maximum and minimum CMF values obtained from the implementation of the GA algorithm for 1000 repetitions of calculations (It = 1000).

Table 5. Evaluation of statistical parameters of the CMF calculation results in GA algorithm for total crashes

Statistical characteristic	Mean	Minimum	Maximum
Value	1.04889	0.70662	1.41518

By evaluating the results presented in Table 4 and comparing with the calculated values of CMF for total crashes, it is determined that the values obtained from the GA evaluation method are fitting and the mean value of CMF for total crashes is about 1. In order to examine the results of CMF calculation based on GA algorithm, cumulative probability curve based on 2000 CMF calculations (1000 times for maximum values and 1000 times for minimum values) is presented (Figure 5). This diagram shows that the probability of CMFs greater than 1 was above 75 percent, so the calculation of CMF values greater than 1 according to the presented results was more than 75 percent reliable.

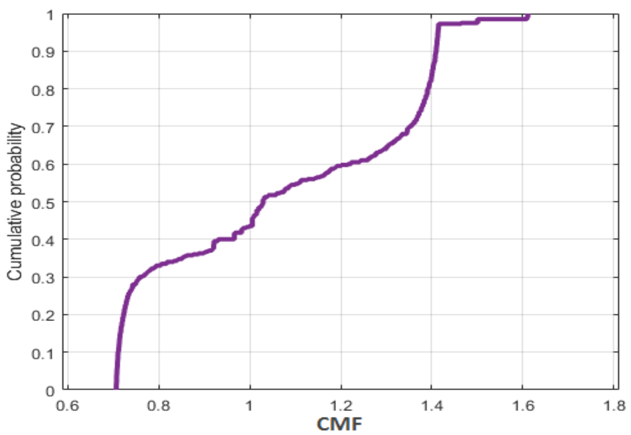

Figure 5. Cumulative probability curves for CMF values calculated by the GA method for total crashes

3.4. Reliability Assessment for Fatal-Injury Crashes

According to the explanations and results presented in the previous section, here the model is executed according to fatal-injury crashes and using GA algorithm, the maximum and minimum CMF values are calculated with 1000 repetitions. The results are presented in Figures 6 and 7 and Table 6.

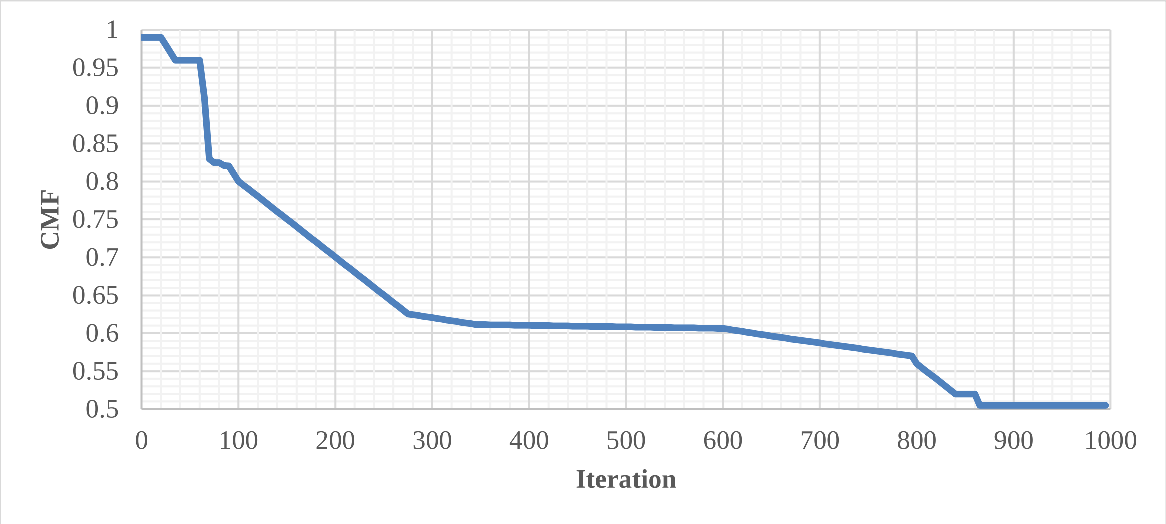

Figure 6. Calculation of minimum CMF values based on GA algorithm for fatal-injury crashes

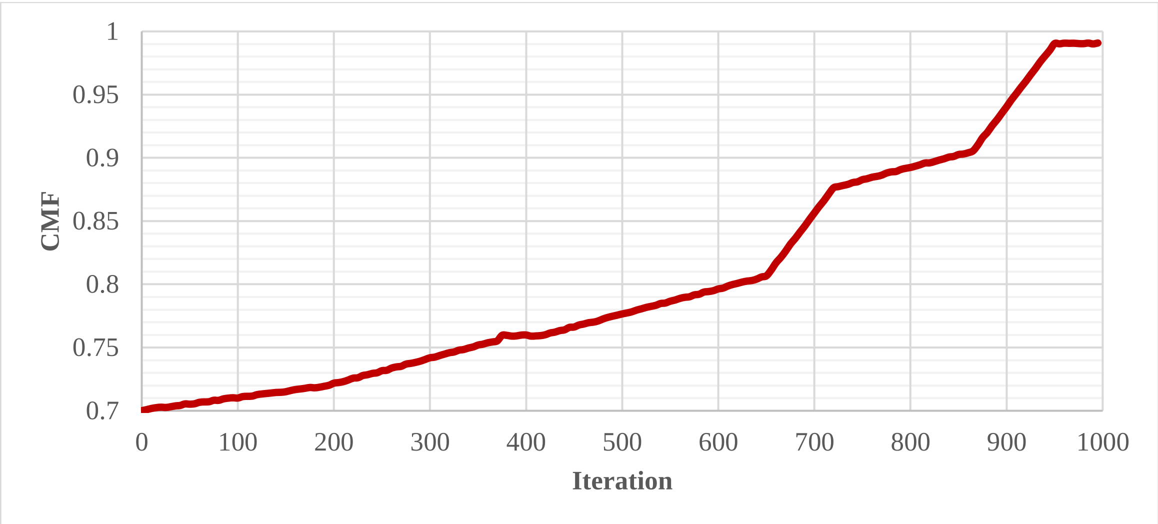

Figure 7. Calculation of maximum CMF values based on GA algorithm for fatal-injury crashes

Table 6. Evaluation of statistical parameters of the CMF calculation results in GA algorithm for fatal-injury crashes

Statistical characteristic	Mean	Minimum	Maximum
Value	0.7078	0.505	0.9919

In order to evaluate the results of CMF calculation based on GA, the cumulative probability curve based on 2000 CMF calculations (1000 times for maximum values and 1000 times for minimum values) is presented.

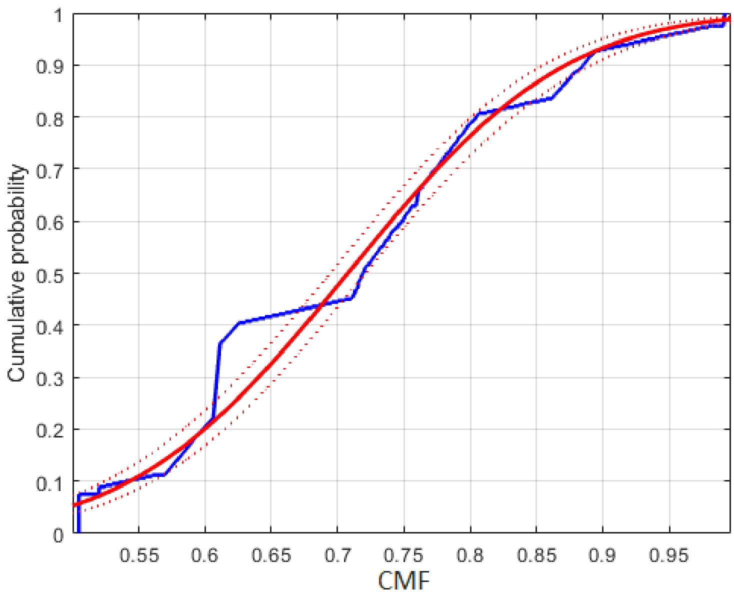

Figure 8. Cumulative probability curves for CMF values calculated by the GA method for fatal-injury crashes

Figure 8 presents the results of the computational values for CMF based on the GA algorithm and for the severity of fatal-injury crashes. Based on the diagram, it is clear that the probability of CMFs less than 1 was more than 95 percent, so the calculation of CMF values less than 1 according to the presented results was more than 95 percent reliable.

3.5. Reliability Ranking

In order to evaluate the quality of CMF values calculated using genetic algorithm, ranking based on three CMF evaluation criteria is proposed in this study. The CMF quality rating system has been based on considering different criteria for evaluating CMFs. In this ranking system, the following three main criteria are considered for CMF rating:

- Evaluation of computational CMF values based on primary data (excluding GA algorithm for calculations): If computational CMF based on primary data has significant values less than 1, this factor has a positive effect and otherwise it has had a negative impact.
- Evaluation of computational CMF values based on GA algorithm: If the mean CMF calculated based on GA algorithm for different types and crash severities is less than 1, it has a positive effect and if it is more than 1, it has a negative effect.
- Evaluation of cumulative probability function for computational CMF values by GA method: According to this criterion, if the values calculated for the CMF have a cumulative probability function greater than 1 with a confidence level of more than 90 percent, it has a negative effect, and if the cumulative probability function has values less than 1 with a confidence level greater than 90 percent, it has had a positive impact.

Based on the defined criteria, eight ranking levels for CMF evaluation are presented in Table 7.

Table 7. Reliability ranking for CMFs

Level	Criterion 1	Criterion 2	Criterion 3	Ranking
1	✓	✓	✓	*****
2	✓	✓	×	****
3	×	✓	✓	****

Level	Criterion 1	Criterion 2	Criterion 3	Ranking
4	✓	×	✓	***
5	×	×	✓	**
6	×	✓	×	**
7	✓	×	×	**
8	×	×	×	*

According to Table 7, the values of the accident adjustment coefficient obtained from the model for two different crash severities (total crashes and fatal-injury crashes) have been evaluated for reliability. Accordingly, the positive and negative effects of each ranking criterion for the two crash severities are presented in Table 8.

Table 8. CMF reliability assessment for the two crash severity levels

Crash severity	Criterion 1	Criterion 2	Criterion 3	Ranking
Total crashes	×	×	✓	**
Fatal-injury crashes	✓	✓	✓	*****

According to the results presented in the previous sections and based on the ranking presented in Table 7, it is clear that the CMF value for remedial action of signalizing intersections with counters, did not have a significant impact on reducing total crashes (CMF=1.04). On the other hand, by evaluating the CMF values calculated for fatal-injury crashes, it is determined that the calculated CMF is approximately equal to 0.75, which indicates the positive effect of the remedial action and reduction of fatal-injury crashes. In addition, according to the proposed rating system, CMF values for fatal-injury crashes have the highest rank, which indicates a very high reliability for the calculated CMF values. Therefore, it is possible to confidently take the remedial action of signalizing intersections with the installation of a counter as an effective measure to reduce the number of fatal-injury crashes.

4. Conclusion

In this study, a model has been defined to evaluate the reliability of Crash Modification Factors (CMFs) using Genetic Algorithm (GA). Also, a ranking system has been proposed to evaluate the CMFs based on the metaheuristic GA method. The model presented in this study is defined by using the before-after study methods and using the Full Bayesian (FB) function, which has a higher computational accuracy than the Empirical Bayesian (EB) method, and is based on the use of comparison group sites. In this model, GA metaheuristic algorithm was used to evaluate the reliability. In this regard, Karaj metropolitan crash data for years 2016-2020 were collected for the practical implementation of the proposed model.

By preparing the initial data, the created model was implemented based on different crash severities and the results were obtained in each stage. In this study, two categories of crashes were evaluated based on severity: total crashes and fatal-injury crashes. The model created in the first stage calculates the values of the CMFs directly using the original primary data. It should be noted that CMF values are calculated based on different conditions using the model parameters (site type, study period, traffic parameters and route length). The results obtained from this section showed that the remedial action of signalizing intersections along with the installation of counters had an average CMF equal to 1.07 for crashes with severity of the first type (i.e. total crashes), which indicates no significant effect of the treatment. In contrast, the results of the assessment of fatal-injury crashes showed that the CMF was 0.75, which indicates the significant effect of corrective action on reducing fatal-injury crashes (which is very important in social costs of road traffic injuries).

In the next step, the reliability of computational CMF values was evaluated using GA algorithm. The results of this step, which was based on statistical parameters and distribution of different computational values of CMF, showed that the CMF for fatal-injury crashes has a higher reliability, so that in this case the CMF value was less than 1 with a reliability above 95 percent. On the other hand, by evaluating the results of the reliability measurement method, it was found that the CMF for total crashes had lower reliability and CMF values of more than 1 resulted in 75 percent reliability. Based on the presented ranking system and evaluation of the modeling results, it was found that the CMF values for 5-star fatal-injury crashes are in the highest rank. In contrast, the CMF of total crashes with 2 stars is ranked lower. By evaluating the CMF values for the two categories of severities including fatal-injury crashes (CMF less than 1, about 0.7) and total crashes (CMF about 1 and equivalent to 1.04) as well as evaluating the reliability obtained based on the ranking system, It is found that the remedial action of installing a traffic signal with a counter will significantly reduce the number of fatal-injury crashes at intersections. In addition, the mentioned remedial action did not have a significant effect on the number of total crashes, which of course, due to the lower reliability needs further analysis.

The proposed model has the ability to evaluate the reliability of CMFs based on the GA algorithm and provide rankings to measure the reliability of computational CMFs in different conditions. It is suggested that while maintaining the overall model presented in the CMF calculation process, other methods such as data mining be used to assess the reliability of CMFs and the results of these methods be compared with the results of this study. For future research, it is suggested that parametric studies be performed on the model presented in this research and that the research model be evaluated under various data and various safety measures. In addition, it is suggested that in order to develop the proposed model, the effect of several remedial actions should be modeled based on the existing model and the results should be compared and evaluated with the results of other models.

Among the limitations of the research, we can mention the selection of a safety measure to increase the accuracy of the research. It is important to combine different corrective measures in accident reduction analyzes and express their effect on each other. The lack of accurate data and information on the types of accidents based on the way the cars collided and the evaluation of the corrective action in reducing the accidents with different types are other limitations of future researches.

It is suggested to define and implement study models based on artificial intelligence methods (machine learning, artificial neural networks) in future researches. According to the characteristics of different artificial intelligence methods, these methods are expected to be very efficient in the evolution of accident study models. In addition, genetic algorithm method has been used in the model structure in the conducted studies to evaluate the reliability of CMF. It is suggested to use other methods such as data mining to evaluate the reliability of CMF by maintaining the generality of the model presented in the CMF calculation process. It is also possible to use the hybrid deep and machine learning model to estimate the traffic volumes of intersections.

5. References

– Abdel-Aty, M.A., Lee, C., Park, J., Wang, J.H., Abuzwidah, M. and Al-Arifi, S., 2014. Validation and application of highway safety manual (part D) in Florida (No. BDK78-977-14). Florida. Dept. of Transportation.

– Ahmed, M.M., Abdel-Aty, M. and Park, J., 2015. Evaluation of the safety effectiveness of the conversion of two-lane roadways to four-lane divided roadways: Bayesian versus empirical Bayes. Transportation research record, 2515(1), pp.41-49.

– Ali, M.S., Kitali, A.E., Kodi, J.H., Alluri, P. and Sando, T., 2021. Safety impacts of transit signal priority using a full Bayesian approach. Transportation research record, 2675(11), pp.1189-1204.

– Al-Marafi, M.N. and Somasundaraswaran, K., 2023. A review of the state-of-the-art methods in estimating crash modification factor (CMF). Transportation Research Interdisciplinary Perspectives, 20, p.100839.

– Amiri, A.M., Sadri, A., Nadimi, N. and Shams, M., 2020. A comparison between artificial neural network and hybrid intelligent genetic algorithm in predicting the severity of fixed object crashes among elderly drivers. Accident Analysis & Prevention, 138, p.105468.

– Tahir, H.B., Yasmin, S., Lord, D. and Haque, M.M., 2023. Examining the performance of engineering treatment evaluation methodologies using the hypothetical treatment and actual treatment settings. Accident Analysis & Prevention, 188, p.107108.

– Brimley, B.K., Saito, M. and Schultz, G.G., 2012. Calibration of Highway Safety Manual safety performance function: development of new models for rural two-lane two-way highways. Transportation research record, 2279(1), pp.82-89.

– Cheng, J. and Li, Q.S., 2008. Reliability analysis of structures using artificial neural network based genetic algorithms. Computer methods in applied mechanics and engineering, 197(45-48), pp.3742-3750.

– El-Basyouny, K. and Sayed, T., 2009. Collision prediction models using multivariate Poisson-lognormal regression. Accident Analysis & Prevention, 41(4), pp.820-828.

– El-Basyouny, K. and Sayed, T., 2010. Full Bayes approach to before-and-after safety evaluation with matched comparisons: case study of stop-sign in-fill program. Transportation research record, 2148(1), pp.1-8.

– El-Basyouny, K. and Sayed, T., 2011. A full Bayes multivariate intervention model with random parameters among matched pairs for before–after safety evaluation. Accident Analysis & Prevention, 43(1), pp.87-94.

– El-Basyouny, K. and Sayed, T., 2012. Measuring safety treatment effects using full Bayes non-linear safety performance intervention functions. Accident Analysis & Prevention, 45, pp.152-163.

– Howlader, M.M., Yasmin, S., Bhaskar, A. and Haque, M.M., 2023. A before-after evaluation of protected right-turn signal phasings by applying Empirical Bayes and Full Bayes approaches with heterogenous count data models. Accident Analysis & Prevention, 179, p.106882.

– Lan, B., Persaud, B., Lyon, C. and Bhim, R., 2009. Validation of a full Bayes methodology for observational before–after road safety studies and application to evaluation of rural signal conversions. Accident Analysis & Prevention, 41(3), pp.574-580.

– Lee, C., Abdel-Aty, M., Park, J. and Wang, J.H., 2015. Development of crash modification factors for changing lane width on roadway segments using generalized nonlinear models. Accident Analysis & Prevention, 76, pp.83-91.

– Li, W., Carriquiry, A., Pawlovich, M. and Welch, T., 2008. The choice of statistical models in road safety countermeasure effectiveness studies in Iowa. Accident Analysis & Prevention, 40(4), pp.1531-1542.

– Lubliner, H. and Schrock, S.D., 2012. Calibration of the Highway Safety Manual prediction method for rural Kansas highways (No. 12-2835).

– Mrowicki, A., Krukowski, M., Turoboś, F., Jaśkiewicz, M., Radkowski, S. and Kubiak, P., 2021. Determining vehicle pre-crash speed in frontal barrier crashes using genetic algorithm model adjustment techniques for intermediate car class. International journal of crashworthiness, pp.1-8.

– Munyazikwiye, B.B., Karimi, H.R. and Robbersmyr, K.G., 2017. Optimization of vehicle-to-vehicle frontal crash model based on measured data using genetic algorithm. IEEE Access, 5, pp.3131-3138.

– National Research Council (US). Transportation Research Board. Task Force on Development of the Highway Safety Manual and Transportation Officials. Joint Task Force on the Highway Safety Manual, 2010. Highway safety manual (Vol. 1). AASHTO.

– Ospina-Mateus, H., Jiménez, L.A.Q., López-Valdés, F.J., Garcia, S.B., Barrero, L.H. and Sana, S.S., 2021. Extraction of decision rules using genetic algorithms and simulated annealing for prediction of severity of traffic accidents by motorcyclists. Journal of ambient intelligence and humanized computing, pp.1-22.

– Park, J., Abdel-Aty, M., Wang, J.H. and Lee, C., 2015. Assessment of safety effects for widening urban roadways in developing crash modification functions using nonlinearizing link functions. Accident Analysis & Prevention, 79, pp.80-87.

– Persaud, B., Lan, B., Lyon, C. and Bhim, R., 2010. Comparison of empirical Bayes and full Bayes approaches for before–after road safety evaluations. Accident Analysis & Prevention, 42(1), pp.38-43.

– Pu, Z., Li, Z., Jiang, Y. and Wang, Y., 2020. Full Bayesian before-after analysis of safety effects of variable speed limit system. IEEE transactions on intelligent transportation systems, 22(2), pp.964-976.

– Sacchi, E. and Sayed, T., 2014. Accounting for heterogeneity among treatment sites and time trends in developing crash modification functions. Accident Analysis & Prevention, 72, pp.116-126.

– Sacchi, E., Sayed, T. and El-Basyouny, K., 2014. Collision modification functions: Incorporating changes over time. Accident Analysis & Prevention, 70, pp.46-54.

– Tun, Y.W., Pedroso, D.M., Scheuermann, A. and Williams, D.J., 2016. Probabilistic reliability analysis of multiple slopes with genetic algorithms. Computers and Geotechnics, 77, pp.68-76.

– Vinayaraj, V.S. and Perumal, V., 2023. Developing Safety Performance Functions and Crash Modification Factors for Urban Roundabouts in Heterogeneous Non-Lane-Based Traffic Conditions. Transportation Research Record, p.03611981231157727.

– Wang, J.H. and Abdel-Aty, M.A., 2014. Comparison of safety evaluation approaches for intersection signalization in Florida (No. 14-0374).

– Wang, J.H., Abdel-Aty, M.A., Park, J., Lee, C. and Kuo, P.F., 2015. Estimating safety performance trends over time for treatments at intersections in Florida. Accident Analysis & Prevention, 80, pp.37-47.