A Hybrid Optimization Framework for Road Traffic Accident Data

Article *in* International Journal of Crashworthiness · December 2019

DOI: 10.1080/13588265.2019.1701905

Bulbula Kumeda
University of Electronic Science and Technology of China

Fengli Zhang
University of Electronic Science and Technology of China

Ghanim M. Alwan
Madenat Al-Elem University College (MAUC)

Forster Owusu
University of Electronic Science and Technology of China

International Journal of Crashworthiness

ISSN: 1358-8265 (Print) 1754-2111 (Online) Journal homepage: https://www.tandfonline.com/loi/tcrs20

A hybrid optimization framework for road traffic accident data

Bulbula Kumeda, Zhang Fengli, Ghanim M. Alwan, Forster Owusu & Sadiq Hussain

To cite this article: Bulbula Kumeda, Zhang Fengli, Ghanim M. Alwan, Forster Owusu & Sadiq Hussain (2019): A hybrid optimization framework for road traffic accident data, International Journal of Crashworthiness, DOI: 10.1080/13588265.2019.1701905

To link to this article: https://doi.org/10.1080/13588265.2019.1701905

Published online: 17 Dec 2019.

INTERNATIONAL JOURNAL OF CRASHWORTHINESS
https://doi.org/10.1080/13588265.2019.1701905

A hybrid optimization framework for road traffic accident data

Bulbula Kumeda[a], Zhang Fengli[a], Ghanim M. Alwan[b], Forster Owusu[a] and Sadiq Hussain[c]

[a]School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu, China; [b]Chemical Engineering Department, Missouri University of Science and Technology, Rolla, MI, USA; [c]Computer Science, Dibrugarh University, Assam, India

ABSTRACT

With the exponential growth of the number of vehicles in the traffic, road traffic accidents have become a fast-growing and wide-spreading menace, causing the loss of precious human life and economic assets. In addition to the fast growth of population and motorization, roads are highly occupied by cars, buses, trucks, motorcycles, minibikes, taxis, pedestrians, animals, and other travelers. The main challenges in the road traffic accident data prediction and analysis is the small size of the dataset that can be used for training. While road traffic accident causes millions of deaths and injuries every year, their density in time and space is –fortunately-low. This makes machine learning very challenging at the local level. The main target of this work is to minimize the Causality Severity of road traffic accidents. Mathematical analysis for the attributes' effect on Causality Severity was guided to select effective twenty decision variables. An optimization search requires a reliable simulated model that depicts the necessary characteristics of the system under study. The good formulation of the simulated model was reduced the non-linearity and interaction between the variables. The simulated model has been verified by successive linear technique with aid of the Successive Linear Programming (SLP). Mixed Integer nonlinear Program (MINLP) has proven a reliable optimization search with the present model. Reasonable agreements have found when compared the simulated solutions with the experimental data. From the output results, we observed that some of the attributes have a positive effect on Causality-Severity, while other attributes have a negative effect on it. In addition to the implementation output of both MINLP and SLP, we can state that the Age of Driver, Speed Limit, and Age of Causality are necessary and sensitive variables for objective Causality Severity. Optimal results would guide for selecting the best conditions.

ARTICLE HISTORY
Received 7 August 2019
Accepted 1 December 2019

KEYWORDS
Road traffic accident; optimization; accident prediction; successive linear programming; mixed integer nonlinear program

1. Introduction

The fast and dramatic growth of population size and the amount of vehicles on the road causes a set of vehicles moving close and slowly in the cities which leads to the crash of vehicles. Presently one of the main transportation issues in the cities is congestion and the main dimensions of urban societies that include traffic systems are mobility. Currently, road traffic injuries are causing a substantial economic crisis to the whole nation-states of the world starting from individuals to their families. This loss mainly includes the cost of medical treatments and the loss of the productive group of society by injury or death, loss of family's income generators, and the members of the family that handle the whole family burden are consumed their time absent from school or work to treat and care for the injured. The accidents not only unfavorably affect the living of individuals but also their family members, as it can lead families into poverty *via* the bearing effects of the episodes.

Every day with the shocking and terrifying consequence of road traffic accidents more than 3,700 people die globally. Furthermore, 1.35 million persons die and 50 million get injured every year on world roads with the involvement of pedestrians, cars, buses, bicycles, motorcycles, or trucks [1]. Pedestrians, motorcyclists, and cyclists take more than half the number of those people who died in road traffic crashes [2]. Reports indicate that people killed by car accidents in the world is greater than those who die with HIV/AIDS [3,4]. This indicates that the existing road safety control measures are not sufficient to alleviate the situation from getting worse.

According to WHO reports, 90% of the road traffic accidents take place in mid and low-income nations while they merely own 60% of the registered cars in the world [5]. This implies that there is an asymmetrical relation between the severity level of the accidents and the number of vehicles in their street. Moreover, people from lower socio-economic backgrounds within developed countries are more probably to be involved in road traffic crashes. From the class of the society, the most suffering age group in a road accident is the young group in which the females are less possibly to be engaged in road traffic crashes compared to the males. In addition to the young group of people with age of under 25 years, the female is again three times less possibly to be killed by traffic accident crashes than males [6].

CONTACT Zhang Fengli fzhang@uestc.edu.cn

In the mid and low developing nations, the growth of road traffic accidents is very high. The death rate because of road traffic crashes in mid and low-income nations is three times greater than high-income nations [5]. According to world health organization reports, traffic injuries place a vast economic crisis in mid and low-income nations. Rendering to the up-to-date existing cost estimates, globally road traffic damages cost $518 billion while in mid and low-income nations the cost is $65 billion which surpasses the whole amount of money that they get in aid for development. The highest portion of pedestrian and cyclist mortalities took place in the Africa continent which is 44% of deaths because of road accident crashes [6,7].

Nowadays road traffic injuries are the top contributor to the death of children and young adults aged 5-29 years old [5,8]. The burden is excessively suffered by pedestrians, motorcyclists, and cyclists, in particular, those living in mid and low-developing nations. To effectively meet any future worldwide target that might be set to save lives, drastic action is needed to put the road improvement safety measures on the ground [9].

Nowadays the transportation system is deeply dependent on the motor vehicle because most of the world population uses it to move from place to place either for business or for entertainment and so on. This an extreme number of automobiles in world road streets are also the main causes of a different range of problems such as environmental pollution to the major problem that is traffic accidents which result in death and injury of human beings. In the mid and low-developing nations, the size of the road infrastructure and the number of motor vehicles are highly unbalanced [9]. The lack of periodical vehicle maintenance and the use of very aged and old cars are also the key causes of a road accident.

The dramatic growth of road traffic accident contributes to the problems of not realizing the accident behavior, components determining the severity of the accidents, and improper handling of the huge data size that's generated from several sources. To solve these problems many research scholars have attempted different research works in the field but there is a big gap in accident severity predictions and discovering the influential causes like season and time in which the accident happened frequently. This contributes to its own impresses for the challenges in the field of analysis and predictions of road accidents. Selecting an appropriate algorithm for accident severity level detection and choosing the right data preprocessing approaches are some of the challenges in the field of study.

There are traditional ways to lessen road traffic accidents. As a consequence, to reduce traffic accidents, it is not always convenient or even needed to have an enormous expense such as creating awareness for the community through promoting the safety rules in use of roads, traffic signs and increasing the infrastructural capacity. Thus, the analytical and predictive techniques are required, such as data mining algorithms in order to make intelligent decisions to avoid further accidents, enhance transportation system and develop some intelligent traffic safety rules. However, most of these approaches concentrate on the existing accident data to detect the causes of traffic accidents and unable to attain the best prediction techniques using a big size and multi-dimensional accident data with general application value. The predictive accuracy and effectiveness of these models are lesser and their results can't adapt to new architectures of edge computing. This motivates us to employ hybrid optimization techniques for minimizing the causality severity rates of the accident with the application of data mining and machine learning tools and techniques to predict and analyze road traffic accident (RTA).

Here in this work, our main contribution is to minimize the Causality severity of an accident which is the main cause of the death of precious human lives and huge loss of economic assets. To overcome this problem we have used a hybrid optimization framework using Successive Linear Programming (SLP) to verify the simulated model successive linear technique and Mixed Integer nonlinear Program (MINLP) which has proven reliable optimization search with the present model. The optimization processed (method) we used is new. In this work, we studied the statistical effect of each attribute on Causality Severity by using mathematical correlation that depended on real data and would observe the coefficient of each variable. Then, we selected reasonable and effective values. The selected method is deterministic and not stochastic (GA). The deterministic approaches are required the analytical properties of the problems while finding their optimal solutions.

2. Organization of the work

This research work has six sections, where section one is the introduction discussed above and this is section two. Section 3 describes the current related works in the area of road traffic accidents using data mining and machine learning methods to solve road accident problems. Section 4 is the methodology part in which we have discussed the data set we used and the pre-processing methods we followed to select the attributes and the flowchart of the proposed model. Section 5 briefly discusses the optimization methods we have used to minimize the causality severity and state the experimental result. The last one is section six, which concludes the work stating the result of the work to prevent road traffic accidents.

3. Literature review

Currently, road traffic accidents have been both developmental and public health concerns and demanded the attention of researchers, civil societies, vehicle companies, governments, and business communities in the whole world. In traffic safety, the most significant and hot research topic is a road traffic accident which gets high attention of the research community in the field. The incidence of a road traffic accident is largely affected by the geometric features of the road, traffic flow, driver's behavior and environmental conditions of the road [10,11].

To investigate and predict the characteristics and frequencies of the accidents from various features and to identify the hazardous locations/hot spot of the accidents, the work in [12] used association rule mining and modularity optimizing community detection approaches. The outcome from the experiment indicates that the community detection algorithm applied in this work is very effective in detecting clusters with visible features.

The work in [13] described accident injury-severity analysis by demonstrating a modeling method that can well realize the distribution of severity level of accidents on expressway sections. Moreover, to analyze the effect of the weather, the traffic and expressway characteristics on this distribution a random parameters (mixed) logit model is estimated. From the analysis result, the model used has significant potential as a practical tool in planning expressway safety.

To predict lane clearance time and accident duration analysis the framework in [14] applied the M5P tree algorithm which has a better performance compared with the traditional prediction algorithms. It is an effective algorithm to deal with missing values, continuous and categorical variables. In this work, the developed M5P model has been compared with the traditional decision tree and regression models and achieved a better prediction performance.

To challenge the problem of vehicle risk prediction through an efficient vehicle prediction model, the authors in [15] proposed the AdaBoost-SO algorithm. In this work, accident risk prediction is done using factual accident data with the application of data analysis and mining techniques. The proposed system provides driving safety assistance and a smart transportation system with a theoretical basis.

In [16] the authors presented a traffic guidance and control technique based on traffic prediction. They used big accident data with the real-time routing algorithm of vehicle dynamic network to provide road pathway for the travelers and to attain a balanced traffic flow. The system can attain the intention of stable traffic flow through the formation of the simulation technique.

To predict a traffic accident risk level efficiently, the work in [17] used big heterogeneous data to learn ranked feature illustration by developing a deep model of Stack denoise Autoencoder. The experimental result depicted that through human mobility, traffic accident risks can be predicted more significantly.

The authors in [18] applied Convolutional Neural Networks to detect the factors that are highly leading to fatal accidents by considering suitable features and efficiently clustering the records. Various features like speed limit, injury severity, time of the accident, drunk driver, month and weather during the accident, human factors, and light conditions are considered to identify the incidence of the accident. Using the given dataset CNN can effectively detect the threat factors that cause a fatal accident. Thus the public can easily avoid accidents by identifying the danger accident regions.

In this [19] work to estimate the nonlinear relationships among crash frequency and the related risk factor, the authors applied an improved RBFNN (radial basis function neural network) by comparing with the traditional Negative Binomial (NB) and Back-Propagation Neural Network (BPNN). The outcome of the experiment showed that the proposed RBFNN model has improved performance than the NB models to identify the relationship between the risk factors and the crash frequency.

The key goal of this [20] work is to address the limitations of classifications methods such as lack of interpretation for humans and their law classification accuracy performance. To address the limitations and to predict the severity of traffic accidents in accordance with the user's preference rather than using a DT, the authors proposed a novel rule-based multi-objective method. In this work to classify and optimize rules according to support, confidence and comprehensibility evaluation metrics, a multi-objective genetic algorithm is customized. According to the classification metrics, the suggested technique has a better performance than the classification methods like SVM, ANN, and conventional DT. Here also in this [21] work they used a hybrid classification and clustering method to explore the crash injury severity of rural roads by comparing the performance of the classification methods earlier and afterward using the clustering.

The research works in [22–26] described different comparative and predictive machine learning methods that are applied for predicting the accident injury severity of motorcycles and motor vehicles. They stated that most of the accidents happen on the roads varies with the types of vehicles and it is essential to study with the consideration of the vehicle types and the susceptible group of the driver in traffic crashes. Motorcyclists are the most susceptible group of drivers in traffic accidents. While other drivers who drive cars and big trucks receive better protection from their vehicles compared with that of motorcyclists.

Road traffic accident prediction is one of the most serious aspects of road safety in which accidents can be predicted before it happens and to take preventive action to avoid it. Currently, the Machine Learning Prediction models are used in varieties of applications from the students' performance prediction [27] to road accident prediction [28].

Here in the literature [29–33] the authors used different statistical and predictive machine learning algorithms through comparisons on the different datasets to classify, evaluate and predict the crash severity of road traffic accidents. They compared and analyze different algorithms and choose the one which has better analysis and prediction performance. They used different evaluation metrics to compare and choose the suitable algorithms for the specified problem.

To optimize and approximate the parameters of the model which can procreate the actual kinematic outcomes from the crash test accurately, the work in [34] used a genetic algorithm. The proposed model in this work enables vehicle companies to have a better vehicle design with a very less crash test. From the result, they determine that it is possible to get the best solutions with a genetic algorithm which can lead to diverse kind of crash speeds.

To identify the causal factors that influence accident severity, the authors in this [35] work employed a novel multi-objective particle swarm optimization (MOPSO)-based partial classification method. Through generating rules that can be measured by several contradictory criteria like comprehensibility and accuracy for each class, the partial classification can deal with the unbalanced dataset. The outcome of this work illustrates that compared to several rule learning algorithms, the proposed model can produce a set of understandable and accurate rules which can show the relationship between accident severity and risk factors.

In this [36] paper, to deal with the prediction of road accidents using the relations achieved from the proposed approach, the authors used the hybrid of Particle Swarm Optimization (PSO)algorithm and Chaos Optimization Algorithms (COA) in order to attain the optimal relationship between the issues involved in the road accidents. Choosing the optimum values for the PSO algorithm parameters from the values generated is the task of Chaos optimization. From the result analysis, the combination of COA and PSO has better predictive accuracy and improved performance compared to the PSO algorithm.

In this [37] work, to predict the traffic accident fatalities efficiently and encourage the symmetrical growth of transport, an accident prediction model based on the support vector machine is demonstrated. The prediction accuracy of the SVM is significantly affected by the selection of parameters. The optimal parameters can be found and the prediction accuracy of the SVM is improved by presenting the swarm optimization. The best optimization parameter is found by introducing an alteration process of the genetic algorithm into particle swarm optimization. In this work using small sample and non-linear data, the traffic accident fatalities are predicted effectively with expected accuracy.

To predict the injury severity of traffic accidents using accident records in Abu Dhabi, the work in [38] proposed an artificial neural network. The authors used K-means to divide the data into three clusters to improve the prediction accuracy of the proposed classification model. The level of injury, which is the predictable variable has four classes i.e. minor, moderate, severe, death. As a relative standard, an ordered probit model is applied in order to validate the predictive performance of the proposed model. The experimental result depicted that the proposed ANN classifier can predict the injury severity of the accident with an expected and reasonable accuracy.

The author in [39] used a k-means clustering analysis and logistic regression to design an accident prediction framework to overcome the problems related to the analysis of large traffic accident data gathered from Seoul. To analyze and process the large size of the data proficiently, they used a Hadoop model and sampling technique to settle the difficult of data Imbalance. Using the classification method the corrected imbalance data are classified into different groups to amend the accuracy of the prediction model.

Deep learning techniques allow different models to learn illustrations of data with numerous levels of generalizations compared to traditional data mining and machine learning methods. This is because the predictive accuracy and effectiveness of these models are lesser and their results can't adapt to new architectures of edge computing [40].

Currently, with the high need for attention in the area of road traffic accidents, few research scholars are attempted to employ various deep learning models to predict the risk and injury severity level of the accident [41–46]. Since deep learning models have the power to discover complex structures from large data and its processing capability, they have been applied and obtained a prominent performance in many areas of the research. While relatively in the area of vehicle transport only a few and concerning topics of traffic congestions flow and accident risk predictions using deep learning are applied [47–49]. To discover the composite interactions between traffic, roadways, environmental components, and traffic crashes, they proposed an enhanced deep learning model.

To discover the complex relations among roadways, an enhanced deep learning model that can learn data representations with numerous levels of abstractions, the authors in [50] applied multivariate negative binomial technique (MVNP). It is implanting into the refinement supervised unit as a regression layer to address the unnoticed heterogeneity matters in the traffic car predictions. To detect a functional network among the descriptive variables and the characteristics illustrations and to perform traffic crash prediction, a feature learning which is unsupervised and a refinement supervised unit are utilized respectively. The outcome result depicts that the models used in this work are effective in predicting traffic crashes.

4. Methodology

4.1. Data pre-processing

In this work, we have used road safety data which contains road traffic injury accidents in the United Kingdom from the year 1979 to 2015. The dataset includes the consequential causalities, the types of vehicles involved in the accident, and the personal accident injury on the public road records reported to the police. We tested the attributes mathematically and selected the effective 20 attributes as shown in Table 1 above. The optimization process we used is new not optimization method (software tool) and it includes several steps:

1. Selecting effective decision variables (attributes), i.e., refining of data. This needs to a mathematical/statistical test.
2. Formulating the Optimization model. Previously using small region technique (dx is small) for solving the low non-linear model. Here our contributions are the preparation of the model by proposed technique (sub-models) that dividing the variables to groups depended on similar nature (action). Each group represents a simulated model. Then, the global model is combined with them. This reduces the interaction and non-linearity in the model before the solving process. So, lessening the

Table 1. Description and symbols of process variables.

Name of attributes/symbol	Description of attributes
Causality_Severity, Y_{OB}	The asperity of the accidents (Fatal, Serious, Slight)
Sex_of_Driver, X_1	The gender of the driver during the accident (Male, Female, Not known,)
Age_of_Driver, X_2	The age of the Driver during the accident (age is given in years, 0-120)
Age-Band_of_Driver, X_3	The age circle of the drivers during the accident (0-5, 6-10, 11-15, 16-20 …)
Car_Passenger, X_4	The car passenger status during the accident (Not a car passenger, Front seat passenger, Rear seat passenger …)
Vehicle_Type, X_5	Vehicle type during the accident (Pedal cycle, Car …)
Age_of_Vehicle, X_6	The service year of the car starting from the manufacturing year (1, 3, 5, 7 …)
Number_of_Vehicles, X_7	The total number of vehicles which take part in the accident (1, 2, 3 …)
Speed_Limit, X_8	The Speed limit of the road during the accident (20, 30, 40, 50 …)
Road_Type, X_9	The type of road during the accident (Roundabout, One-way street …)
Light_Conditions, X_{10}	A light condition during the accident (Daylight, Darkness-lights lit …)
Weather_Conditions, X_{11}	Condition of the weather during the accident (Fine no high winds, Raining no high winds …)
Road_Surface_Conditions, X_{12}	The surface condition of the road during the accident (Dry, wet/Damp … ..)
Urban_or_Rural_Area, X_{13}	The place of the accident (Urban, Rural, Unallocated)
Causality_Class, X_{14}	The causality of the accident (Driver or rider, Pedestrian, Passenger …)
Sex_of_Causality, X_{15}	The gender of the causality during the accident (Male or Female)
Age_of_Causality, X_{16}	The age of the causality during the accident (age is given in years, 0-120)
Age_Band_of_Causality, X_{17}	The age circle of the Causality during the accident (0-5, 6-10, 11-15, 16-20 …)
Causality_Type, X_{18}	The type of Causality during the accident (Pedestrian, Cyclist …)
Pedestrian_Location, X_{19}	The exact location of the pedestrian during the accident (Not a Pedestrian, Crossing on pedestrian crossing facility …)
Pedestrian_Movement, X_{20}	The movement of the pedestrian during the accident (Not a Pedestrian, Crossing from the driver's nearside …)

number of iteration and time of computing and obtain reliable results as explained below.

3. Selecting of the optimization method. This depends on the experience and use of proper software tools.

In this work, we studied the statistical effect of each attribute on Causality Severity by using mathematical correlation that depended on real data and would observe the coefficient of each variable. Then, we selected reasonable and effective values. The selected method is deterministic and not stochastic (GA). The deterministic approaches are required the analytical properties of the problems while finding their optimal solutions. The flowchart of the proposed model is explained well in Figure 1 below.

4.2. Successive linear programming (SLP)

SLP is an optimization method used to solve nonlinear optimization problems nearly *via* a sequence of linear programs. The technique is established on solving the sequence of first-order approximation i.e. linearization of the model starting at some of the optimal solutions. The linear programming problems are that can be solved effectively which is linearization. Trust regions or related methods are necessitated to confirm intersections in the hypothesis, as the linearization need not be bounded [51].

We are still considering the constrained nonlinear programming problem:

$$\begin{array}{ll} Minimize & f(x) \\ Subjectto, & g_j(x) \le 0\, j = 1, 2, \ldots, m \\ & h_k(x) = 0\, k = 1, 2, \ldots, p \\ & X_i^l \le x_i \le x_i^u\, i = 1, 2, \ldots, x_n \\ & X \in X_1, X_2, \ldots, X_n \end{array}$$

In the SLP method, by solving a sequence of linear programming problems, the answer for the nonlinear programming problem can be found. The attractive option here is to evade the complications related to nonlinearities using linearization. We must know well the problem that we are going to optimize in advance by recognizing how to linearize it.

4.3. Mixed-Integer nonlinear optimization (MINLP)

MINLP is the part of optimization that deals with nonlinear problems with continuous and integer variables. With the object function and/or the constraints, it deals with the problems of continuous, discrete variables and nonlinear functions. It has proven that it's a powerful tool for modeling and chain algorithmic design contests from nonlinear and combinatorial optimization [52].

The general form of MINLP is

$$\begin{gathered} \min f(x, y) \\ s.t. c_i(x, y) = 0\, \forall\, i \in E \\ c_i(x, y) \le 0\, \forall\, i \in I \\ x \in X \\ y \in Y integer \end{gathered}$$

Where each $c_i(x, y)$ is a mapping from R^n to R, E and I are index sets for equality and inequality constraints, respectively. Typically, the functions f and c_i have some smoothness properties, i.e. once or twice continuously differentiable.

5. Modeling and optimization

Optimization is the study to discover the values of decision variables that match and afford the minimum or maximum of one or more wanted targets or real functions by consistently selecting the values of integer or real variables. Manufacturing and engineering activities will not be as effective as they are now, without the optimization of design and operations [53]. Robust optimization could determine the uncertainty variables of the process. For complex nonlinear processes, the simple and effective model may be difficult to derive. The researcher should always cautiously

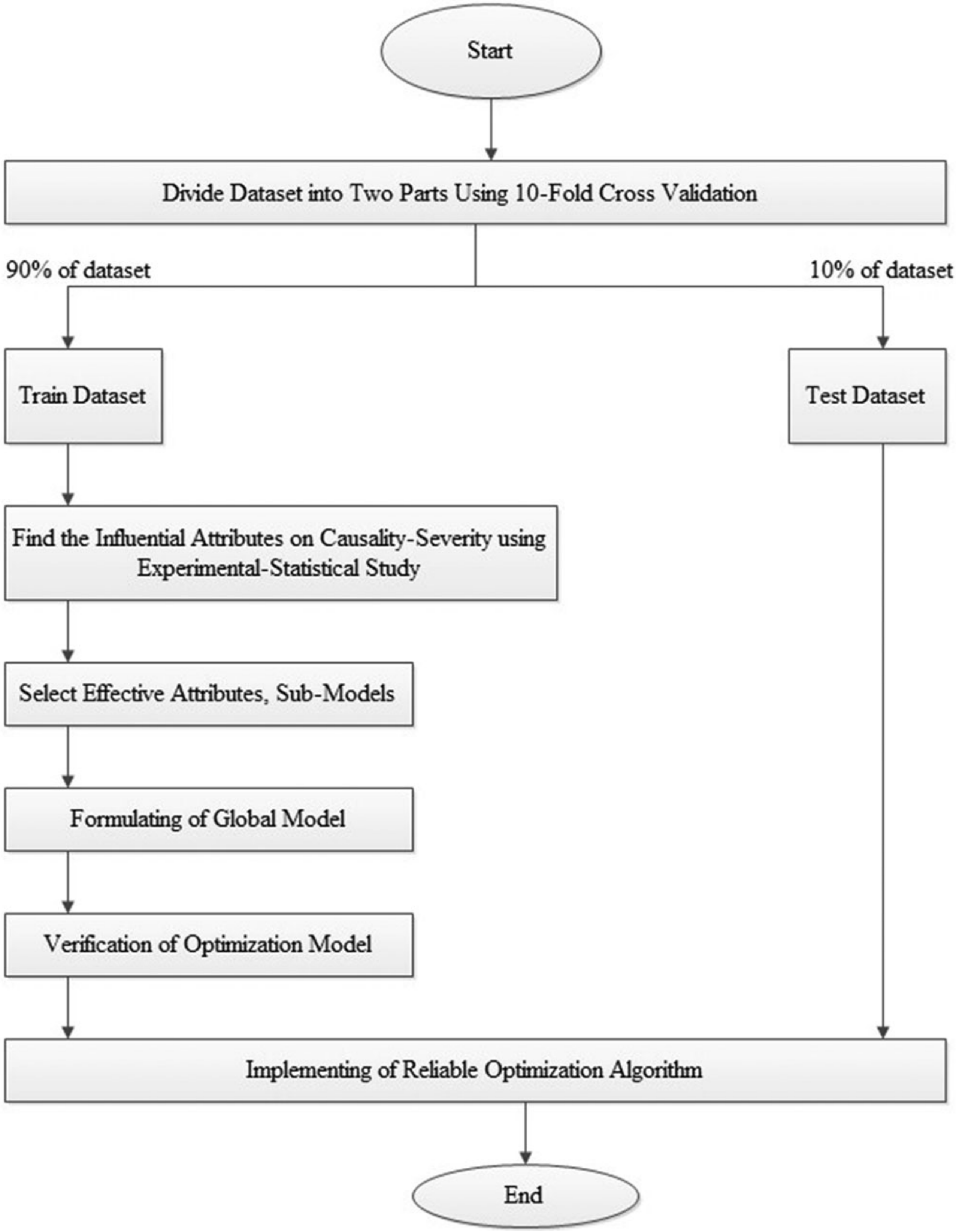

Figure 1. Flow chart of the proposed model.

and carefully formalize the engineering model. Optimization of an imprecise model is modestly enlightening at best and deceptive and a waste of time at worst. Frequently optimization algorithms will feat faintness in the model if they exist.

To approximately reflect the true goal of the optimization problem, attention must be given to optimizing with respect to the objective function [54]. In our work, we make sure that the optimization problem symbolizes the real problem that we need to figure out.

5.1. Formulating of optimization model

Designing an optimization model of a physical problem is the most significant and vital thing in the progress of solving the optimization problem. It requires very good know-how together with the sympathy of the physical process itself. Data collection, problem definition and formulation, model development, model validation and assessment of performance and application and explanation of model are the steps required to model and build an optimization problem [55].

The optimization search requires a reliable simulated model that depicts the essential features of the procedure under study. In general, an optimization model will comprise of the following three items [56]:

- Objective Function: is an expression that states exactly what you want to optimize. It is a mathematical formula that we want to minimize or maximize subject to certain constraints.
- Variables: the quantities which you have under your control are known as variables. It is important and essential to decide the important values from the variables. Because of this reason, variables are sometimes known as decision variables. To discover the values of a model's variable that produce the best value for the objective function, subject to any restraining condition placed on the variable is the key goal of the optimization.
- Constraints: there will be some restriction in values of the variable where a model can accept at least one resource will be restricted almost without exception (e.g., raw materials, time, the budget of your department, etc.). This restriction can be uttered in terms of formulas that are the purpose of the model's variables. Since they restrain the values of the variables can take, these formulas are known to as constraints.

Table 2. Output results of the SLP algorithm.

Y_{OB}	3.00
X_1	1.00
X_2	10.00
X_3	4.00
X_4	0.00
X_5	1.00
X_6	0.00
X_7	1.00
X_8	20.00
X_9	3.00
X_{10}	1.00
X_{11}	1.00
X_{12}	1.00
X_{13}	1.00
X_{14}	1.00
X_{15}	1.00
X_{16}	10.00
X_{17}	2.00
X_{18}	0.00
X_{19}	0.00
X_{20}	0.00

In this work, the main objective is to minimize Causality_Severity. The selected attributes which are affecting the objective are twenty decision variables as shown in Table 1 above. For reliability, it is better to formulate objective function empirically to facilitate the optimization process. We used the real data that were collected at different conditions. Because of a large number of attributes, the sub-models technique will be used. This technique correlates the attributes with the Causality Severity by four equations depended on their similar nature as shown in Table 1 above.

It has proved that the effect factor of each equation on the global model are:

$$Y_1\ (X_1 - X_4)\ \text{is 25\%},\ Y_2\ (X_5 - X_7)\ \text{is 20\%},$$
$$Y_3\ (X_9 - X_{13})\ \text{is 45\%, and}\ Y_4\ (X_{14} - X_{20})\ \text{is 10\%}.$$

This technique will reduce the non-linearity and interaction of variables of the optimization model. So, the number of iteration and time of computing will be dropped.

The optimization model was built depending on the correlation of the collected experimental data. The steps of model building are 1. Refining of experimental data, 2. Limiting of objectives and decision (effective) variables, 3. Formulating of the model by statistical deterministic algorithm, and 4. Validation of the model (by comparison between the simulated results from a model with real

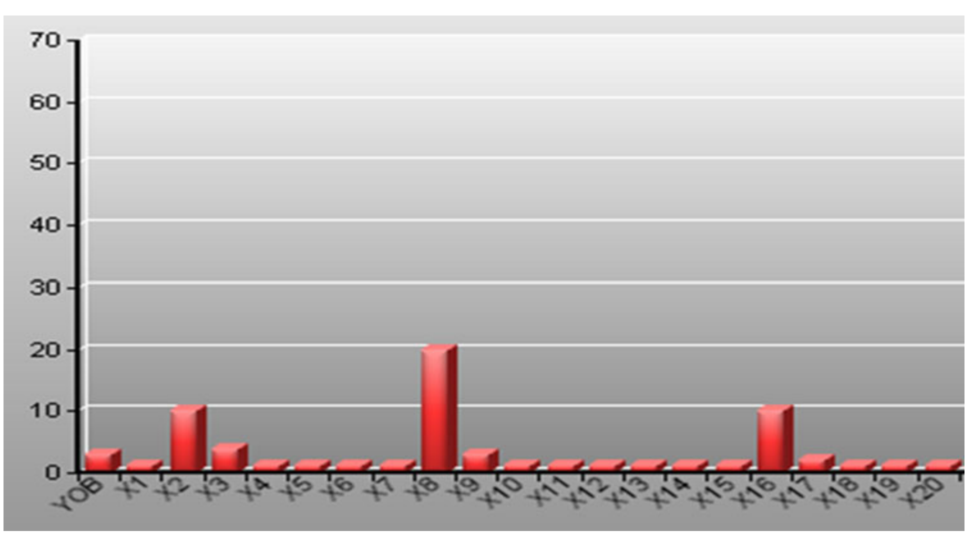

Figure 2. Histogram of simulated results.

experimental data within boundary conditions). The software tool (Statistica Version 10) was implemented to obtain a reliable mathematical structure. In this work, the advanced nonlinear least squares model estimation (Levenberg Marquardt method) was implemented with the aid of the software tool (Statistica Version 10). The Levenberg-Marquardt (LM) algorithm is an enhancement of the classic Gauss-Newton technique for solving nonlinear least-squares regression problems. Statistica data miner is an advanced analytics software platform that includes data mining, machine learning, data management, statistics, and text analytics and data visualization processes as well as a variety of, classification, clustering, exploratory techniques, and predictive modeling. Statistica data miner is the most effective, complete system, and user-friendly tool for the whole data mining procedure - from querying to producing the final reports.

The modified sub-correlation equations are:

$$Y_1 = 2.8712 * (X_1)^{0.011-0.003} * (X_2) + 0.015 * (X_3) + 0.033 * (X_4) + 0.0004 * (X_2) * (X_3) \quad (1)$$

$$Y_2 = 2.83 * (X_5)^{0.0098-0.00198} * (X_6) + 0.053 * (X_7) - 0.002 * (X_8) + 0.00003 * (X_5) * (X_7) \quad (2)$$

$$Y_3 = 3.021 * (X_9)^{-0.0087-0.0088} * (X_{10}) + 2.435 * (X_{11}) - 2.436 * (X_{12}) + 0.068 * (X_{13}) \quad (3)$$

$$Y_4 = 0.53 * (X_{14})^{2.95+0.507} * (X_{15})^{2.92-0.001} * (X_{16}) + 0.534 * (X_{17}) + 0.021 * (X_{18}) - 0.013 * (X_{19}) + 0.01 * (X_{20}) \quad (4)$$

The level of confidence is 95.0% (alpha = 0.050).

Therefore, the global weighted equation represents the combination of sub-equations 1–4:

$$\begin{aligned} Y_{OB} = {} & 0.25 * (2.8712 * (X_1)^{0.011-0.003} * (X_2) + 0.015 * (X_3) \\ & + 0.033 * (X_4) + 0.0004 * (X_2) * (X_3)) \\ & + 0.2 * (2.83 * (X_5)^{0.0098-0.00198} * (X_6) \\ & + 0.053 * (X_7) - 0.002 * (X_8) + 0.00003 * (X_5) * (X_7)) \\ & + 0.45 * (3.021 * (X_9)^{-0.0087-0.0088} * (X_{10}) \\ & + 2.435 * (X_{11}) - 2.436 * (X_{12}) + 0.068 * (X_{13})) \\ & + 0.10 * (0.53 * (X_{14})^{2.95+0.507} * (X_{15})^{2.92-0.001} * (X_{16}) \\ & + 0.534 * (X_{17}) + 0.021 * (X_{18}) - 0.013 * (X_{19}) + 0.01 * (X_{20})). \end{aligned} \quad (5)$$

Boundary Conditions:

$$\begin{gathered} X_1 <= 3, X_1 >= 1; X_2 >= 10, X_2 <= 70; X_3 <= 10, X_3 >= 4; \\ X_4 >= 0, X_4 <= 2; X_5 <= 20, X_5 >= 1; X_6 >= -1, X_6 <= 16; \\ X_7 <= 3, X_7 >= 1; X_8 >= 20, X_8 <= 40; X_9 <= 6, X_9 >= 3, \\ X_{10} >= 1, X_{10} >= 7; X_{11} <= 4, X_{11} >= 1; X_{12} >= 1, X_{12} <= 2; \\ X_{13} >= 1, X_{13} <= 2; X_{14} <= 3, X_{14} >= 1; X_{15} >= 1, X_{15} <= 2; \\ X_{16} <= 51, X_{16} >= 10; X_{17} >= 2, X_{17} <= 8; X_{18} >= 0, \\ X_{18} <= 9; X_{19} >= 0, X_{19} <= 6; X_{20} >= 0, X_{20} <= 9. \end{gathered} \quad (6)$$

The simulated global model (Eq. 5) shows the non-linearity and interaction between the variables (attributes) are low. Also; some of the attributes have a positive effect on objective Causality-Severity, while others attributes have a negative effect on it.

5.2. Verification of optimization model

In this work, we have used a software tool (LINGO18) to solve the global simulated model (Eq. 5) that is subjected to boundary conditions (Eq. 6). LINGO is a complete tool designed to make constructing and solving nonlinear optimization models easier, faster, and more proficient. Depending on the form and specification of (Eq. 5), the successive linear programming (SLP) algorithm has implemented. This technique uses a linear numerical approximation of the true nonlinear model within successive, small regions. The software of the simulated model was designed with the best style. The roots of (Eq. 5) are explained in **Table 2**. The reliability of the global model has been proven that the simulated values stay in the acceptable limits (Eq. 6). **Figure 2** illustrates the histogram of the obtained variables by the SLP algorithm of the non-linear model. Response and sensitivity of the attributes to Causality-Severity are explained in **Figure 3**. **Figures 2** and **3** indicate that X_2, X_8, and X_{16} are effective and sensitive variables for Y_{OB}.

5.3. Optimization search

The optimization technique was a powerful tool to guide the experimental work for desired operating conditions. The main target of solving an optimization model is to find the optimal values of the decision variables.

In this work, the main objective is to minimize Causality Severity that has been limited between 1.0 to 3 .0. Hence, for the optimization process, (Eq. 5) shall modify to a new form:

$$\textbf{Minimize}\{\ Y_{OB}\ \} \quad (7)$$

Subject *to Inequality constraints of (Eq. 6)*

Several deterministic methods were implemented to solve the optimization problem (Eq. 7). The optimization process depends mainly on first-degree derivative and second-degree derivative of (Eq. 7) subjected to (Eq. 6). The selected Mixed Integer nonlinear Program (MINLP) has been proven a reliable search to obtain integer optimal values. The enhancement of optimization results was observed due to the good formulating of the simulated model and selecting the proper software. **Table 3** illustrates the outputs of the algorithm's solutions. The objective (Y_{OB}) is minimized to the average value of the accepted range. **Figure 4** explains the histogram of optimal attributes that have values within the limit of inequality constraints (Eq. 6). The variables X_2, X_8, and X_{16} are still the effective and sensitive variables for Y_{OB} as shown in **Figures 4** and **5**. **Figure 6** shows the feasibility zone of optimal attributes. The Radar Chart shows 20 attributes with its values (red lines). The optimal values are

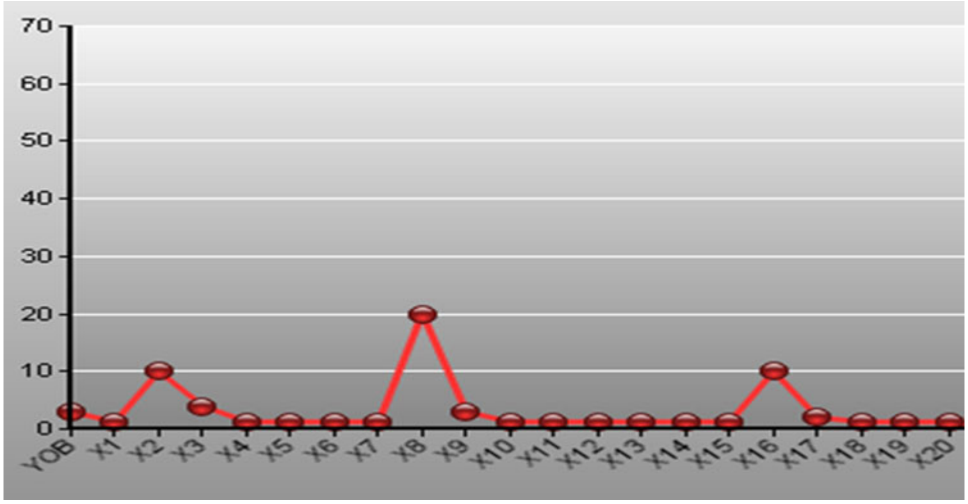

Figure 3. The response of attributes to causality severity.

Table 3. Optimal results of the MINLP algorithm.

Y_{OB}	2.00
X_1	1.00
X_2	70.00
X_3	4.00
X_4	0.00
X_5	1.00
X_6	16.00
X_7	1.00
X_8	40.00
X_9	6.00
X_{10}	7.00
X_{11}	1.00
X_{12}	2.00
X_{13}	1.00
X_{14}	1.00
X_{15}	1.00
X_{16}	51.00
X_{17}	2.00
X_{18}	0.00
X_{19}	6.00
X_{20}	0.00

surrounded by 41 constraints that have been generated by the MINLP search process.

6. Conclusions

With the advent of vehicle technology and road infrastructural development, the mobility of people from place to place increased in an exponential way. To prevent and control road traffic injuries, it needs the commitments of all the community, governmental and non-governmental sectors to take actions that address the welfare of roads, vehicles, and road users in an all-inclusive manner. This can be accomplished through effective involvements in designing innocuous infrastructure and integrating road safety characteristics into transport planning and land-use, refining the safety characteristics of vehicles, taming post-crash care for victims of road crashes, making and implementing laws relating to

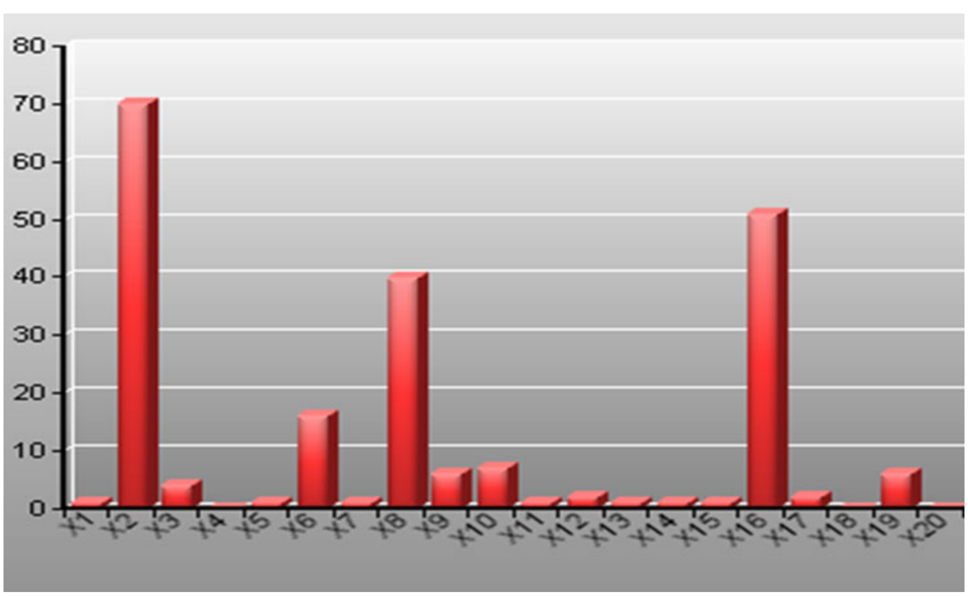

Figure 4. Histogram of optimal attributes at minimum causality severity.

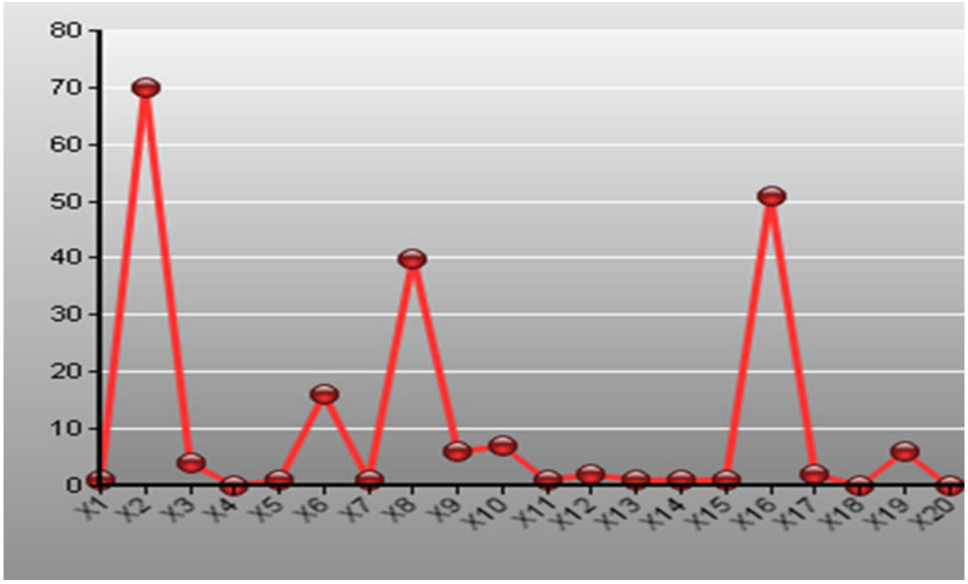

Figure 5. The response of attributes at optimum condition.

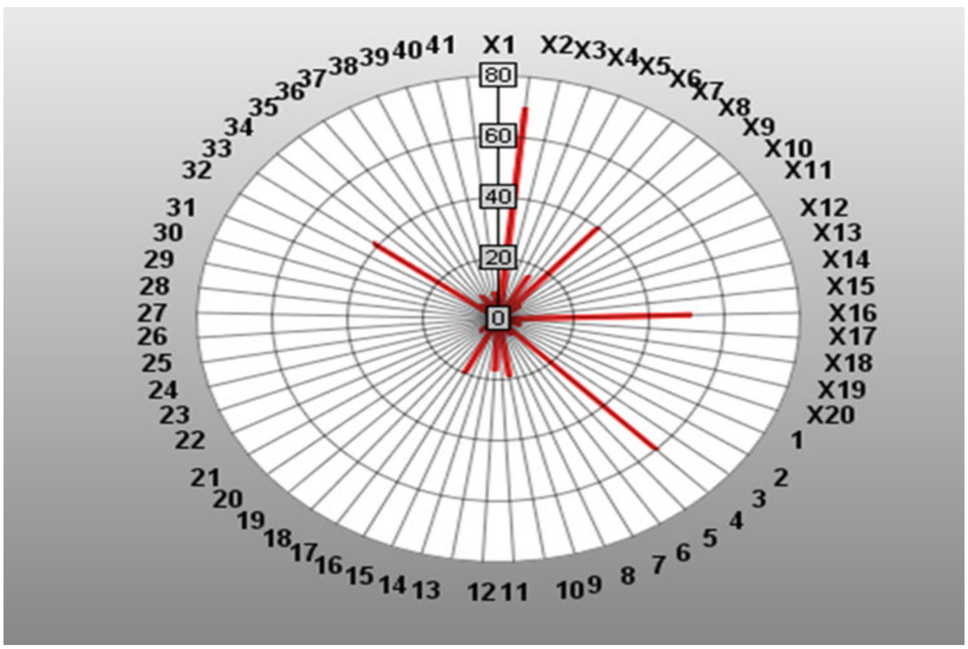

Figure 6. Radar chart of optimal attributes.

main risks, and increasing public awareness. The most dangerous risk factors that lead to accidents are: - Over speed, driving under the influence of psychoactive materials and alcohol, non-use of motorcycle helmets, seatbelts and child constraints, unfocussed driving, unsafe vehicle, risky road infrastructure, inadequate post-crash care, insufficient implementation of traffic laws.

Recently the major national health concern in the world is road traffic accidents (RTA) which bring about the death of priceless human lives and loss of a huge amount of assets. Numerous factors such as a human, environmental, vehicle, etc. are highly associated with traffic accidents, some of those issues are more important in defining the severity of the accident than others. To predict the most significant factors among environmental, human and vehicle, data mining analytical solutions are significantly applied to determine the severity of RTA's. Most of the communal factors that are associated with traffic accidents are the infrastructural factors which include the environmental and road infrastructure, the human factors that are mainly related to the features of the road users and the features of the vehicles. Moreover in deciding the severity of the accident some of the factors are more significant than others. Therefore, to disclose more patterns and knowledge that can be employed in prevention and safety, the analysis of the determining factors of accident severity is very apparent.

In this research work, the simulated model was represented successfully by sub-models equations. The optimization process would enhance by the good formulation of the optimization model, selection of effective decision variables and a proper design for software optimization problems. In this work, we have used Statistica software Version 10 for formulating optimization model based on advanced nonlinear estimation. The good formulation of the simulated model was reduced the non-linearity and interaction between the variables. LINGO is a powerful tool for solving modeling and optimization problems. The successive linear algorithm (SLP) has been found as the best solver for the true nonlinear model within successive, small regions. The Mixed-Integer Nonlinear Programming (MINLP) algorithm has been proven the reliable deterministic search for the global nonlinear optimization problem to obtain integer values. Optimal results would guide for selecting the best-operating conditions. This can lessen the menace of experimental runs and the cost is consumed for operating. A reasonable agreement has been found when compared the simulated results with the experimental data. From the output/results, the optimization search could indicate an accuracy of 95%. From the implementation output of both SLP and MINLP algorithms, we can state that Age_of_Driver, Speed_Limit, and Age_of_Causality are effective and sensitive variables for Causality_Severity (Y_{OB}).

Disclosure statement

No potential conflict of interest was reported by the authors.

References

[1] World Health Organization, et al. Road traffic injuries. World Health Organization; 2018. Available from: http://www.who.int/mediacentre/factsheets/fs358/en/

[2] Road traffic injuries among vulnerable road users [Internet]. [cited 2018. Nov 11]. Available from: http://www.euro.who.int/data/assets/pdf_file/0004/98779/polbrief_road_injuries.pdf

[3] World Health Organization. Global health expenditure Database [Internet]. 2018. http://apps.who.int/nha/database

[4] World Health Organization. Global health estimates [Internet]. World Health Organization; 2018. Available from: http://www.who.int/healthinfo/global_burden_disease/en/

[5] Global status report on road safety 2018: summary. Geneva: World Health Organization; 2018. (WHO/NMH/NVI/18.20) License: CC BY-NC-SA 3.0 IGO).

[6] Wang SC. Treat the patient, not just their disease.In: Proceedings of the 2017 International IRCOBI Conference on the Biomechanics of Injury. September 13-15, 2017; Antwerpm, Belgium.

[7] Nantulya VM, Reich MR. The neglected epidemic: road traffic injuries in developing countries. BMJ. 2002;324(7346): 1139–1141.

[8] World Health Organization. Disease, injury, and causes of death country estimates, 2000–2015 [Internet]. World Health Organization; 2017. [cited 2018 Oct 29]. Available from: http://www.who. int/healthinfo/global_burden_disease/estimates_country_2000_2015/en/

[9] United Nations General Assembly. 64/255. Improving global road safety [Internet]. 2010. [cited 2018 May 16]. Available from: http://www.who.int/violence_injury_prevention/publications/road_traffic/

[10] Li JL, Sun WW. Cause analysis of traffic accidents on express highway and study on their countermeasures [J]. China Saf Sci J. 2005;15(1):59–62.

[11] Liu Q, Lu H, Zhang Y, et al. Characteristic analysis and countermeasure study on road traffic accidents in China. China Saf Sci J. 2006;16(6):123–128.

[12] Lin L, Wang Q, Sadek AW. 2014. Data mining and complex network algorithms for traffic accident analysis. Transportation Research Board 93 rd Annual Meeting (No. 14-4172).

[13] Milton JC, Shankar VN, Mannering FL. Highway accident severities and the mixed logit model: an exploratory empirical analysis. Acc Anal Prev. 2008;40(1):260–266.

[14] Zhan C, Gan A, Hadi M. Prediction of lane clearance time of freeway incidents using the m5p tree algorithm. IEEE Trans Intell Transport Syst. 2011;12(4):1549–1557.

[15] Zhao H, Yu H, Li D, et al. Vehicle accident risk prediction based on AdaBoost-so in vanets. IEEE Access. 2019;7: 14–549–14–557.

[16] Xin C, Na C, Yeshuai B. Analysis on key technologies of traffic prediction and path guidance in intelligent transportation. 2016 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS). IEEE; 2016. p. 5–8.

[17] Chen Q, Song X, Yamada H, et al. Learning deep representation from big and heterogeneous data for traffic accident inference. 13th AAAI Conference on Artificial Intelligence; February 12 –17, 2016, Phoenix, Arizona, USA.

[18] Yasaswini L, Mahesh G, Siva Shankar R, et al. Identifying road accidents severity using convolutional neural networks. IJCSE. 2018;6(7):354.

[19] Huang H, Zeng Q, Pei X, et al. Predicting crash frequency using an optimized radial basis function neural network model. Transportmetrica A: Transp Sci. 2016;12(4):330–345.

[20] Hashmienejad SH-A, Hasheminejad SMH. Traffic accident severity prediction using a novel multi-objective genetic algorithm. Int J Crashworthiness. 2017;22(4):425–440.

[21] Hasheminejad SH-A, Zahedi M, Hasheminejad SMH. A hybrid clustering and classification approach for predicting crash injury severity on rural roads. Int J Injury Control Saf Promotion. 2018;25(1):85–101.

[22] Wahab L, Jiang H. A comparative study on machine learning-based algorithms for prediction of motorcycle crash severity. PloS One. 2019;14(4):e0214966.

[23] Jeong H, Jang Y, Bowman PJ, et al. Classification of motor vehicle crash injury severity: a hybrid approach for imbalanced data. Acc Anal Prev. 2018;120:250–261.

[24] Cigdem A, Ozden C. Predicting the severity of motor vehicle accident injuries in Adana-turkey using machine learning methods and detailed meteorological data. Int J Intell Syst Appl Eng. 2018;6(1):72–79.

[25] Wahab L, Jiang H. Severity prediction of a motorcycle crash with machine learning methods. Int J Crashworthiness. Vol. 24, 2019:1–8.

[26] Chung Y, Song T-J, Yoon B-J. Injury severity in delivery-motorcycle to vehicle crashes in the Seoul metropolitan area. Acc Anal Prev. 2014;62:79–86.

[27] Hussain S, Muhsion ZF, Salal YK, et al. Prediction model on student performance based on internal assessment using deep learning. Int J Emerg Technol Learn. 2019;14(8):4–22.

[28] Hussain S, Muhammad L, Ishaq F, et al. Performance evaluation of various data mining algorithms on road traffic accident dataset. In: Information and communication technology for intelligent systems.Smart Innovation, Systems and Technologies, Springer, vol 106. 2019, p. 67–78, Singapore.

[29] Kunt MM, Aghayan I, Noii N. Prediction for traffic accident severity: comparing the artificial neural network, genetic algorithm, combined genetic algorithm and pattern search methods. Transport. 2012;26(4):353–366.

[30] Singh J, Singh G, Singh P, et al. Evaluation and classification of road accidents using machine learning techniques. In: Emerging research in computing, information, communication, and applications. Advances in Intelligent Systems and Computing, vol 882. 2019, p 193-204, Springer, Singapore.

[31] García Cuenca L, Puertas Sanz E, Aliane N, et al. Traffic accidents classification and injury severity prediction. In 2018 3rd IEEE International Conference on Intelligent Transportation Engineering (ICITE) (pp. 52–57). IEEE.

[32] Iranitalab A, Khattak A. Comparison of four statistical and machine learning methods for crash severity prediction. Acc Anal Prev. 2017;108:27–36.

[33] Ahmadi A, Jahangiri A, Berardi V, et al. Crash severity analysis of rear-end crashes in California using statistical and machine learning classification methods. J Transp Saf Secur. Vol. 11, 2018:1–25.

[34] Munyazikwiye BB, Karimi HR, Robbersmyr KG. Optimization of vehicle-to-vehicle frontal crash model based on measured data using a genetic algorithm. IEEE Access. 2017;5:3131–3138.

[35] Qiu C, Wang C, Fang B, et al. A multiobjective particle swarm optimization-based partial classification for accident severity analysis. Appl Artif Intell. 2014;28(6):555–576.

[36] Gharehchopogh FS, Dizaji ZA. A new chaos agent-based approach in prediction of the road accidents with a hybrid of pso optimization and chaos optimization algorithms: a case study. IJAR. 2014;6(2):108.

[37] Gu X, Li T, Wang Y, et al. Traffic fatalities prediction using support vector machine with hybrid particle swarm optimization. J Algorithms Comput Technol. 2018;12(1):20–29.

[38] Alkheder S, Taamneh M, Taamneh S. Severity prediction of traffic accident using an artificial neural network. J Forecast. 2017;36(1):100–108.

[39] Park S-h, Kim S-M, Ha Y-G. Highway traffic accident prediction using VDS big data analysis. J Supercomput. 2016;72(7): 2815–2831.

[40] LeCun Y, Bengio Y, Hinton G. Deep learning. Nat Int J Sci. 2015;521(7553):436–444.

[41] Sameen MI, Pradhan B, Shafri H, et al. Applications of deep learning in severity prediction of traffic accidents. Global Civil Engineering Conference. Springer 2017, vol 9. p, 793–808, Singapore.

[42] Ren H, Song Y, Wang J, et al. A deep learning approach to the citywide traffic accident risk prediction. 21st International Conference on Intelligent Transportation Systems (ITSC). IEEE; 2018. p. 3346–3351, Maui, United States.

[43] Sameen M, Pradhan B. Severity prediction of traffic accidents with recurrent neural networks. Appl Sci. 2017;7(6):476.

[44] Zheng M, Li T, Zhu R, et al. Traffic accident's severity prediction: a deep-learning approach-based CNN network. IEEE Access. 2019;7:39–897–39 910.

[45] An J, Fu L, Hu M, et al. A novel fuzzy-based convolutional neural network method to traffic flow prediction with uncertain traffic accident information. IEEE Access. 2019;7:20708–20722.

[46] Zhang Z, He Q, Gao J, et al. A deep learning approach for detecting traffic accidents from social media data. Transp Res Part C: Emerging Technol. 2018;86:580–596.

[47] Huang W, Song G, Hong H, et al. Deep architecture for traffic flow prediction: deep belief networks with multitask learning. IEEE Trans Intell Transport Syst. 2014;15(5):2191–2201.

[48] Lv Y, Duan Y, Kang W, et al. Traffic flow prediction with big data: a deep learning approach. IEEE Trans Intell Transp Syst. 2014;16(2):865–873.

[49] Polson NG, Sokolov VO. Deep learning for short-term traffic flow prediction. Transp Res Part C: Emerging Technol. 2017; 79:1–17.

[50] Dong C, Shao C, Li J, et al. An improved deep learning model for traffic crash prediction. J Adv Transp. 2018;2018:1.

[51] Belsnes MM, Wolfgang O, Follestad T, et al. Applying successive linear programming for stochastic short-term hydropower optimization. Electr Power Syst Res. 2016;130:167–180.

[52] Kılınç MR, Sahinidis NV. Exploiting integrality in the global optimization of mixed-integer nonlinear programming problems with BARON. Optim Methods Software. 2018;33(3): 540–562.

[53] Rangaiah GP. Multi-objective optimization: techniques and applications in chemical engineering. Vol. 1. World Scientific, 2009.

[54] Parkinson AR, Balling R, Hedengren JD. Optimization methods for engineering design. Brigham Young University 2013;5:11.

[55] Eren Y, KU¸CUkdemiral IB, Anď ΓU. Introduction to optimization. In: Optimization in renewable energy systems.recent perspectives. ButterworthHeinemann, Elsevier 6th March 2017, p. 27–74.

[56] Huang G-B, Ding X, Zhou H. Optimization method based extreme learning machine for classification. Neurocomputing. 2010;74(1–3):155–163.