**FACTA UNIVERSITATIS**
Series: **Mechanical Engineering**
https://doi.org/10.22190/FUME241128018H

**Original scientific paper**

# ENHANCING SUPPORT VECTOR MACHINE PERFORMANCE IN FORECASTING THE NUMBER OF VEHICLES INVOLVED IN TRAFFIC CRASHES VIA METAHEURISTIC OPTIMIZATION ALGORITHMS

**Sina S. Haghshenas[1], Mohammad Hassan M. Seraji[1], Sami S. Haghshenas[1], Giuseppe Guido[1], Vittorio Astarita[1], Vladimir Simic[2,3,4], Dragan Marinkovic[5]**

[1]Department of Civil Engineering, University of Calabria, Via Bucci, Rende, Italy
[2]University of Belgrade, Faculty of Transport and Traffic Engineering, Belgrade, Serbia
[3]Yuan Ze University, College of Engineering,
Department of Industrial Engineering and Management, Taoyuan City, Taiwan
[4]Department of Computer Science and Engineering, College of Informatics,
Korea University, Seoul, Republic of Korea
[5]Institute of Mechanical Science, Vilnius Gediminas Technical University, Lithuania

**Abstract**. *Traffic accidents cause huge financial, human, and emotional losses every year. With increased urbanization and population growth, safety problems are also increasing. This study develops a novel approach by integrating Support Vector Machine (SVM) with advanced meta-heuristic algorithms to build a classification model of the number of vehicles involved in a traffic accident and the ranking of factors contributing to the accident, which is essential for improving transportation safety management. For this purpose, nine factors contributing to impacting road safe transport in urban areas of Cosenza, southern Italy, are assessed and ranked through the application of a stochastic approach. These features were determined using machine learning techniques, including SVM, combined with some metaheuristic optimization methods involving the Cuckoo Search Algorithm (CSA), Multi-verse Optimizer (MVO), and Whale Optimization Algorithm (WOA). With the SVM-WOA algorithm offering a great accuracy value, the results reveal that these approaches are successful in detecting elements influencing traffic safety management in road transportation.*

**Key words**: *Traffic Safety, Accident Factors, Urban areas, Machine Learning, Metaheuristic Optimization Methods*

Received: November 28, 2024 / Accepted April 15, 2025
**Corresponding author**: Vladimir Simic
University of Belgrade, Faculty of Transport and Traffic Engineering, Vojvode Stepe, Belgrade, Serbia
E-mail: vsima@sf.bg.ac.rs

## 1. INTRODUCTION

Traffic accidents pose a severe challenge globally in terms of great economic losses, human fatalities, injuries, and even serious emotional and psychological effects on people and society. Growing urbanization and population seem to increase the need for ensuring better road safety and adopting effective traffic management strategies [1, 2].

Basically, improvement in road safety relates to the identification and understanding of various factors leading to traffic accidents. The value of such knowledge is quite substantial for decisions and steps transportation professionals need to take while designing and implementing appropriate measures to minimize the risk of accidents and make roads safer [1, 3]. Conventional methods for traffic accident prediction, most of them rooted in statistical models and regression analysis, normally operate under the assumption of linear relationships between contributing factors and accident occurrences. These models assume linear relationships, which often fail to capture the nonlinear dynamics of real-world traffic systems. These limitations have consequently made advanced computational methods, including techniques of artificial intelligence (AI) and machine learning (ML), increasingly embraced by researchers in their struggle to improve predictive modeling capability [4]. The nature of complexity in such approaches means they are very good at unraveling complex patterns and relationships that may exist yet remain obscure in high-dimensional data, an attribute particularly appropriate in the context of traffic accident prediction.

Among these, support vector machines (SVMs) have drawn significant attention due to their capability to deal with nonlinear and high-dimensional input effectively. SVMs are grounded in statistical learning theory. They also perform with a great deal of consistency compared to conventional statistical models that not only stress better accuracy but also generalized capability. Other researchers have focused on improving the prediction capability of the SVMs using metaheuristic optimization algorithms [5]. The metaheuristic algorithms are inspired by natural phenomena such as evolutionary processes and swarm behavior that provide ways of improving the parameters used by the SVMs with the aim of increasing their predictive accuracy and efficiency. A wide range of metaheuristic methods, each with distinct advantages and constraints, have been effectively used to optimize SVM models in several areas, including the crucial space of traffic accident prediction [6-8].

This is notwithstanding the fact that all these highly remarkable achievements have been recorded over the recent years in the area of traffic accident prediction using SVMs combined with metaheuristic optimization. Besides, many challenges persist. First, the computational complexity of developing processes for the SVM models makes it difficult to deploy them in real situations and further scalability is hard, especially when united with metaheuristic optimization [5]. Second, the volume and quality of the input data, including real-time traffic flow and weather, fundamentally influence the accuracy of prediction models [9]. Consequently, this implies that, with increasing urgency for finding a solution to such problems, computation-efficient algorithms that can emerge patterns from big quantities of data in real-time and provide predictions with real timescales are in need [10].

This study addresses known limitations in traffic accident prediction research and offers a multifarious contribution. At first, the research improves the ability to predict by combining SVMs with a number of metaheuristic optimization techniques. By systematically comparing various algorithms, their pros and cons are made clear, making it easy to pick the most appropriate optimization method for any given traffic situation.

The study also looks at urban roads in Cosenza, which is in southern Italy, in a way that is specific to that area. This shows how local factors can affect predicting accidents. As a result, the most important factors that can cause accidents can be found and ranked, meeting the need for safety solutions that are tailored to the specific situation.

Lastly, this study fills in a gap in the existing body of work by applying new developments in theoretical modeling to real-world problems that have to do with scalability and high-dimensional computational complexity. This means that there are now even better algorithms that can handle large amounts of traffic data while still being accurate. These algorithms can give policymakers and transportation engineers useful information that they can use to make the roads safer.

## 2. LITERATURE REVIEW

From traditional statistical methods to state-of-the-art AI systems, this literature review investigates the development of methodologies in traffic accident analysis and prediction, therefore stressing important trends and possible future directions of study. The application of conventional methods, such as traditional statistics and economic methods, has proved helpful in understanding traffic. Sipos et al. [11] performed a spatial econometric analysis, especially spatial autocorrelation and geographically weighted regression, to find the geographical trend of traffic accidents. This would show how the spatial factors affect the probability of occurrence of accidents. Lord and Mannering [12] conducted a thorough study and evaluation of the different statistical methods used to look at crash-frequency data. This showed how important it was to choose the right methods for each situation. Vedagiri and Killi [13] used surrogate safety measures like time-to-collision and post-encroachment time in analyzing uncontrolled intersections in mixed traffic streams. This illustrates how the more general approaches may be used to evaluate traffic safety across particular situations. Also, Sawalha and Sayed [14] looked at it from the temporal point of view. They applied regression analysis on crash data from major arterial roads to develop how safety varies over time and the impacts of actions on that variation in safety changes. Traditional statistical methods can be applied under various circumstances, as shown by Dell'Acqua and Russo's [15] work in developing the safety performance function for low-volume roads by negative binomial regression. All these functions take into account the difficulties arising from fluctuating road and traffic conditions.

The introduction of data mining and ML has opened up new possibilities for forecasting and analyzing traffic accidents. Li et al. [16] used data mining techniques such as association rule mining and decision trees to determine the causes of one category of pointless catastrophic traffic accidents, thereby demonstrating the usefulness of this type of method in processing crash data. Pourroostaei Ardakani et al. [17] elaborated further on this by showing the practicality of ML techniques with the use of models like Random Forest (RF) to carry out predictions on road car accidents. They underlined how these techniques raised forecast accuracy, which would help to produce more accurate and efficient traffic safety rules. Wang et al. [18] proposed a more advanced strategy based on multi-view multi-task spatiotemporal networks, incorporating convolutional neural networks and graph convolutional networks, where various kinds of data and various tasks are effectively combined for a global prediction.

Researchers have also concentrated on improving predicting models and handling natural data difficulties in traffic accident research. Guido et al. [19] studied feature selection methods including information gain and correlation-based feature selection as well as screening the possible factors regarding safety management, indicating their contribution to the improvement of predictive accuracy. Advanced data modeling, such as that using, for example, the SARIMA model of Deretić et al. [20], has made it possible to forecast traffic accidents based on historical patterns. Dimitrijevic et al. [21] used the RF-based classifier to include both proactive and reactive accident data in their models as well as improve the accuracy and location-specific forecasts of short-term crash risk prediction. Danesh et al. [22] developed data leveling methods including random under-sampling and SMOTE and applied metaheuristic optimization approaches, including the Genetic Algorithm, to guarantee unbiased model performance, thereby addressing the issue of imbalanced datasets in crash prediction.

A growing area of study in recent years is also the combination of traditional statistical methods with AI techniques. The hybrid model proposed by Dong et al. [23] combined a state-space framework with support vector regression. This model showcased the potential of merging statistical and ML techniques to enhance the prediction of traffic accidents. The model took into consideration both temporal and spatial dynamics present in the data. Mannering and Bhat [24] provided a review of the analytic methods in accident research, including conventional and upcoming techniques. They also provided a glimpse of future directions for the same. The practical implementation of ML in real-world applications is reflected in a case study carried out by Sufian et al. [25] on traffic accident prediction using heterogeneous urban data, where they employed various ML models and showed their practical implementation.

## 3. METHODOLOGY

Traffic accident prediction and factor analysis involve navigating complex, often non-linear relationships within high-dimensional datasets. While traditional statistical methods are indeed useful, they capture the underlying complexity rather inadequately. So, in this study, the power of ML is used; i.e. SVMs that have high recognition for handling nonlinear and high-dimensional data with great efficiency. To enhance the predictive capabilities of SVMs, we couple them with nature-inspired metaheuristic optimization techniques to tune model parameters for efficient performance optimization.

The theoretical underpinning of SVM and the adopted metaheuristic algorithms in this paper, namely, the Cuckoo Search Algorithm (CSA), Multi-Verse Optimizer (MVO), and Whale Optimization Algorithm (WOA), will be included in the following sections. This also encompasses the methodological framework related to data preparation, model construction, and evaluation. The approach, with its integration of SVMs and metaheuristic optimization, shall provide a robust and accurate means toward the prediction of the number of vehicles involved in a traffic accident and the identification of those significant factors that determine such accidents for the development of appropriate road safety measures.

### 3.1. Support Vector Machine

Artificial intelligence (AI) and ML approaches are being used by many academics and industry sectors, which are changing how we approach complex problems and enhancing operational efficiency [26-29]. SVM is a successful ML algorithm developed by Cortes

and Vapnik in 1995 [30]. It is a supervised learning algorithm used in regression and classification tasks. SVMs are linear classifiers that maximize the margin between two classes by creating a classification hyperplane in the center of the maximum margin. The objective is to identify the optimal hyperplane in an n-dimensional space. Two labels are assigned for classification: +1 for cases above the hyperplane and -1 for those below. Eq. (1) displays a collection of sample sets utilized in the process of learning data for categorization [31]:

$$S = \{(x_i, y_i)_{i=1}^{n} | x_i \in R^N, y_i \in \{-1,1\}, i = 1,2,\ldots\ldots, l\} \tag{1}$$

Here, $x_i$ marks the data and $y_i$ is the target variable for the *i*-th sample.

Determined by applying current support vectors and constraints, the ideal hyperplane is the one with the biggest margin among the produced hyperplanes. Eqs. (2) and (3) express constraints [32]:

$$Min \frac{1}{2} \|w\|^2 \tag{2}$$

$$s.t.\ y_i(wx_i + b) \geq 1 \tag{3}$$

where *w* denotes the weight vector and *b* denotes the bias vector [30].

Subsequently, after taking into account an error coefficient, the constraints are reformulated and rectified based on Eqs. (4) and (5). The purpose of this error coefficient is to enhance the precision of categorization [33]:

$$Min \frac{1}{2} \|w\|^2 + c \sum_{i=1}^{n} \varepsilon_i \qquad (\varepsilon_i \geq 0) \tag{4}$$

$$s.t. \begin{cases} y_i(wx_i + b) \geq 1 - \varepsilon_i \\ c \geq 0 \end{cases} \qquad (i = 1,2,3,\ldots\ldots, n) \tag{5}$$

The penalty coefficient is denoted by the symbol *c*. Next, the Lagrange approach is employed to address classification difficulties in SVM. These problems are formulated as a dual optimization problem, which is based on Eq. (6) [34]:

$$\begin{cases} W(a) = \sum_{i=1}^{n} a_i - \frac{1}{2} \sum_{i,j=1}^{n} a_i\, a_j y_i y_j K(x_i, x_j) \\ s.t. \sum_{i=1}^{n} a_i y_i (0 \leq a_i \leq c; i = 1,2,3,\ldots., n) \end{cases} \tag{6}$$

The function *K* is a mathematical function referred to as the kernel function. Various types of kernel functions are available including the linear (LIN), radial basis function (RBF), and polynomial (POL). Parameters gamma ($\gamma$) and *d* define the several kinds of kernels for RBF and POL. Here, *d* denotes the degree term of the polynomial (POL) kernel function, while gamma ($\gamma$) is used for both RBF and polynomial kernels [35].

The main responsibility of the kernel is to convert the input data into the needed structure. The accuracy of categorization could change depending on the knowledge about the application of several kernel functions in pertinent situations. Eqs. (7-9) provide respectively the formulas of certain kernel functions, including LIN, RBF, and POL [36]:

$$G(x_i, x_j) = x_i^t x_j \tag{7}$$

$$G(x_i, x_j) = exp(-\gamma \left\| x_i - x_j \right\|^2) \tag{8}$$

$$G(x_i, x_j) = (-\gamma x_i^t x + 1)^d \tag{9}$$

### 3.2. Whale Optimization Algorithm

The metaheuristics algorithms are considered very important in the area of engineering, especially because many practical problems require complex system and process optimization for which other methods have proven inefficient [37-40]. The Whale Optimization Algorithm (WOA) is a metaheuristic algorithm that uses swarm behavior to solve continuous optimization challenges. It is straightforward to develop and robust, comparable to nature-inspired algorithms [41]. The WOA model involves the humpback whale population in a complex search space, with individual whale positions as choice variables and distance between them and food as the objective cost. The movement of an individual whale is determined through three processes: encircling prey, spiral bubble-net feeding technique, and prey search. This method is comparable to nature-inspired algorithms due to its simplicity and robustness [42].

Humpback whales surround their prey after sensing where it is. The WOA method supposes the target prey, or in close approach to the optimal solution, is the current best candidate solution. While the remaining search agents change their locations to be closer to the ideal search agent, it aims to identify the best search agent. Mathematically, this behavior is expressed by the following equations [41]:

$$\vec{D} = \left|\vec{C}.\overrightarrow{X^*}(t) - \vec{X}(t)\right| \tag{10}$$

$$\vec{X}(t+1) = \overrightarrow{X^*}(t) - \vec{A}.\vec{D} \tag{11}$$

$$\vec{A} = 2\vec{a}.\vec{r} - \vec{a} \tag{12}$$

$$\vec{C} = 2 \cdot \vec{r} \tag{13}$$

where $\overrightarrow{X^*}$ denotes the current best position, $\vec{X}$ represents the position of the current whale, $t$ indicates the recent iteration, $\vec{A}$ and $\vec{C}$ are coefficient vectors, $\vec{a}$ represents a linearly decreasing value ranging from 2 to 0 during the iterations, and $\vec{r}$ is a random number uniformly distributed between 0 and 1 [40]. The symbol ‘‘| |’’ denotes the absolute value, and element-wise multiplication is shown by “·” [41].

A mathematical equation is used to simulate the spiral movement pattern observed between humpback whales and their prey, resembling the helix-shaped motion of these whales [42]:

$$\vec{X}(t+1) = \overrightarrow{D'}e^{b1}.\cos(2\pi l) + \overrightarrow{X^*}(t) \tag{14}$$

$$\vec{X}(t+1) = \begin{cases} \overrightarrow{X^*}(t) - \vec{A}.\vec{D} & if\ p < 0.5 \\ \overrightarrow{D'}e^{b1}.\cos(2\pi l) + \overrightarrow{X^*}(t) & if\ p \geq 0.5 \end{cases} \tag{15}$$

where $\overrightarrow{D'}$ represents the distance of the ith whale to the prey, $b$ denotes a constant defining the spiral shape, $l$ is a random number in [–1, 1], whereas $p$ is a randomly chosen value that falls within the range of -1 to 1 [42].

In the exploration phase of prey search, global optimizers are utilized. If the value of $A$ is greater than 1 or less than -1, the search agent is updated according to a randomly selected search agent, replacing the current best search agent [42]:

$$\vec{D} = \left|\vec{C}.\overrightarrow{X_{rand}} - \vec{X}\right| \tag{16}$$

$$\vec{X}(t+1) = \overrightarrow{X_{rand}} - \vec{X}.\vec{D} \tag{17}$$

where $\overrightarrow{X_{rand}}$ is selected arbitrarily from the whales in the current iteration [42].

### 3.3. Cuckoo Search Algorithm

The Cuckoo Search Algorithm (CSA) is a metaheuristic optimization technique inspired by the brood parasitism behavior observed in certain cuckoo species. In nature, some cuckoos lay their eggs in the nests of other host birds, relying on the host to raise their offspring. However, there is a probability that the host bird may detect and discard the foreign eggs [43]. Three basic guidelines defined by CSA help to replicate cuckoo behavior [44]:

1. Each cuckoo lays one egg in a randomly chosen host nest; the number of available nests remains constant.
2. High-quality eggs will be reserved for the next generation.
3. Host birds run the chance of seeing and discarding alien eggs.

A constant switch parameter $P_a$ balances the global search (exploration) and local search (exploitation) that define CSA. Levy fly [44] is proposed to implement it for the global exploration:

$$x_i^{t+1} = x_i^t + \alpha \oplus \text{Levy}(s, \lambda) \tag{18}$$

where $\alpha$ is the scale factor, $\oplus$ denotes entry-wise multiplications, $\lambda$ ($1 < \lambda < 3$) is the power coefficient, Levy $(s, \lambda)$ is the characteristic scale, which may be computed as $x_i^t$ and $x_i^{t+1}$, respectively.

$$\text{Levy}(s, \lambda) \approx \frac{\lambda\Gamma(\lambda).\sin(\pi\lambda/2)}{\pi} \frac{1}{s^{1+\lambda}}, \qquad (s \gg s_0 \gg 0) \tag{19}$$

Formulated in Eq. (19), $\Gamma$ symbolizes the gamma function, $s_0$ is the initial step size of Levy flight, and $s$ represents the step size of Levy flight [45]:

$$s = \frac{U}{|V|^{1/\lambda}} \tag{20}$$

$U$ and $V$ denote random values following normal distribution:

$$U \sim N(0, \sigma^2), V \sim N(0,1) \tag{21}$$

where $\sigma$ is the variance of the normal distribution followed by $U$, which is computed as follows [45]:

$$\sigma^2 = \left[\frac{\Gamma(1+\lambda)}{\lambda\Gamma((1+\lambda)/2)} \cdot \frac{\sin(\pi\lambda/2)}{2^{(\lambda-1)/2}}\right]^{1/\lambda} \tag{22}$$

One may arrange the local search for CSA as follows [45]:

$$x_i^{t+1} = \begin{cases} x_i^t + \alpha s \otimes H(p_a - \epsilon) \oplus \left(x_j^t - x_k^t\right), r > Pa \\ x_i^t \, , otherwise \end{cases} \tag{23}$$

where $H(p_a - \epsilon)$ is the Heaviside function, $\epsilon$ is a random integer taken from the uniform distribution, where $x_j^t$ and $x_k^t$ are two solutions acquired via random permutation.

### 3.4. Multi-Verse Optimization

MVO is a metaheuristic technique exploring search space by simulating wormholes, black holes, and white holes. An inflation rate between exploitation and exploration balances wormhole to black-white hole ratios. In this context, candidate solutions are universes; decision variables are things or components found within those universes. The inflation rate controls the likelihood of finding both white (exploration) and black holes (exploitation); i.e. low rates indicate the best chance of a black hole and high rates indicate the highest possibilities. A roulette wheel selection mechanism is used in MVO to solve the analytical model and exchange the items between the worlds of black and white hole tunnels. The solution space exhibits a random global representation as [46]:

$$U = \begin{bmatrix} x_1^1 & x_1^2 & \cdots & x_1^d \\ x_2^1 & x_2^2 & \cdots & x_2^d \\ \vdots & \vdots & \vdots & \vdots \\ x_n^1 & x_n^2 & \dots & x_n^d \end{bmatrix} \tag{24}$$

In this context, *U* symbolizes the world, n denotes the frequency of search components, *d* represents the measurements of control parameters, and $x_i^j$ represents the jth parameter of the i-th world, which follows the form described below [46]:

$$x_i^j = \begin{cases} x_k^j & r1 < NI(Ui) \\ x_i^j & r1 \geq NI(Ui) \end{cases} \tag{25}$$

where *Ui* represents the ith world, *NI* represents the normalized inflation rate, $r_1$ represents a random integer between 0 and 1, and $x_k^j$ represents the jth variable of the kth world, selected using the roulette wheel scenario.

The operational mechanism of the transferred items via wormholes is shown as follows [46]:

$$x_i^j = \begin{cases} \begin{cases} x_j + TDR \cdot \left[ \left( ub_j - \right) \cdot r_4 + lb_j \right] & r_3 < 0.5 \\ x_j - TDR \cdot \left[ \left( ub_j - \right) \cdot r_4 + lb_j \right] & r_3 \geq 0.5 \end{cases} & r_2 < WEP \\ x_i^j & r_2 \geq WEP \end{cases} \tag{26}$$

here $x_j$ is the jth variable of the optimum world; *WEP* (Wormhole Existence Probability) and *TDR* (Traveling Distance Rate) are coefficients; $ub_j$ and $lb_j$ indicate the upper and lower limit of jth variable, $x_i^j$ indicates the jth parameter of the ith world; $r_4$, $r_3$, and $r_2$ are accidental numbers belonging to [0, 1]. *WEP* takes the following form and enhances the exploitation [46]:

$$x_i^j = WEP = min + t \times \left( \frac{max - min}{T_{max}} \right) \tag{27}$$

where the variable *i* represents the current repetition, $T_{max}$ represents the greatest repetition frequency, and "*min*" and "*max*" represent the lowest and maximum values of the controlled variables. The acquisition of *TDR* is done in the following manner [46]:

$$TDR = 1 - \frac{t^{(1/p)}}{T_{max}^{(1/p)}} \tag{28}$$

where, $p$ represents the level of precision achieved during the iterations of exploitation. As the value of $p$ increases, the exploitation/local search process becomes faster and more precise.

## 4. Case Study

The proposed method for analyzing road accidents in Italy was tested using a dataset of 564 accident records from the Regional Road Accident Center (CRISC) of the Calabria region, covering the years 2001-2020. CRISC acquires on its platform the accidents recorded by ISTAT in the regional territory. ISTAT, the Italian National Institute of Statistics, is the primary source of official statistical data in Italy, responsible for gathering and producing information about the Italian economy and society. The data is gathered through a monthly survey involving various authorities, such as traffic police, carabinieri, and municipal police, who use the ISTAT CTT/INC model known as "Road Accidents" to document each accident involving a vehicle on the road network that resulted in injuries [47].

The survey parameters include road accidents, fatalities, and injuries. Road accidents involve accidents on a road open to public traffic, resulting in injuries or fatalities to one or more individuals and involving at least one vehicle. Fatalities include individuals who perish immediately or within 24 hours following the accident, while injuries are categorized as serious or minor injuries [47].

A comprehensive total of data on accident accidents were meticulously documented and assessed across urban and rural areas of the Calabria region, in southern Italy (Fig. 1). These accidents have been categorized and grouped based on several criteria (Table 1).

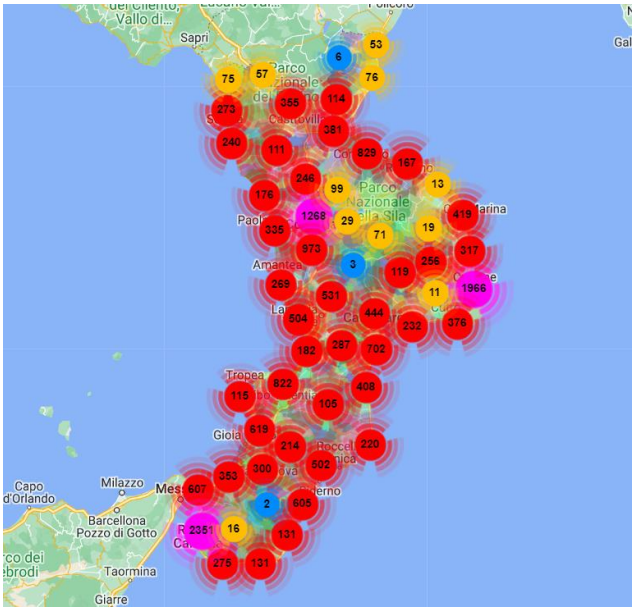



**Fig. 1** Accident sites in the Calabria region for the years 2001-2020. (Source: CRISC Regione Calabria)

**Table 1** Accident database fields were analyzed

| Data field type | Data field | Description |
|---|---|---|
| Human characteristic | Driver Gender | Male or Female |
| | Driver Age | Young, Middle-age, Old |
| Vehicle characteristic | Vehicle type | Motorcycle, Car, Bus, Truck, Other |
| Road | Road type | Two carriageways, More than two carriageways, One two-way carriageway, One one-way carriageway; |
| | Road signal | Absent, Horizontal, Temporary construction, Vertical, Vertical and horizontal; |
| | Road surface | Dry, Wet, Icy, Snow, covered, Slippery; |
| Environment | Date light | Daylight and Nighttime |
| | Weather conditions | Clear/Sunny, Hail, Rain, Snow, Fog, Strong wind, Other; |
| Location | Urban area | Intersection, Non-intersection, Others; |
| Accident characteristic | Number of vehicles | Number of vehicles involved, one or more than one. |

## 5. MODELLING PROCESS

This study used a binary classification method to identify patterns and correlations between different input elements and the number of vehicles involved in accidents. The objective of the binary classification model was to determine whether any correlation between daylight, location, road type, road surface, road signs, weather conditions, types of vehicles, driver age, and driver gender against the number of vehicles as the dependent variable. Identifying the control parameters and performance indices are the major steps of algorithms to reach the optimal binary model [48].

In this regard, the current study developed and compared four machine-learning algorithms; i.e. SVM, SVM-CSA, SVM-WOA, and SVM-MVO. The input dataset consisted of nine factors related to the number of vehicles in the urban area of Cosenza, southern Italy. Data were classified by two classes: '1', in the case when at most one car participated in an accident, and '2' in the case with at least two or more involved vehicles. The main class separation criterion considered was based on taking into account the minimum number of vehicles involved in an accident.

The best classification models were developed and compared, and the sensitivity analysis was carried out to determine the relevance of an effect independently exerted by each factor. The confusion matrix's accuracy and error in binary classification modeling are practical performance indicators. Data normalization is essential for data-driven system modeling methods, as greater scale factors might lead to computational deviations. All data undergoes min-max normalization prior to modeling.

### 5.1. Correlation Analysis

This study highlights that computing and controlling the parametric correlation of independent input datasets are of prime importance in providing correct results and to avoid misinterpretation of the data. Pearson's correlation coefficient, sometimes also referred to as the bilateral correlation coefficient or Pearson product-moment correlation coefficient, is one of the most useful methods to assess linear connections between variables. Eqs. (29–32) depict the relationships of Pearson's correlation coefficient [49].

$$\rho = r = \frac{SP_{Dxy}}{\sqrt{SS_X.SS_Y}} \tag{29}$$

$$SP_{Dxy} = \sum xy - \frac{(\sum x)(\sum y)}{n} \tag{30}$$

$$SS_X = \sum_i^n x_i^2 - \frac{(\sum x_i)^2}{n} \tag{31}$$

$$SS_Y = \sum_i^n y_i^2 - \frac{(\sum y_i)^2}{n} \tag{32}$$

Here, $X$ and $Y$ constitute the independent parameters; $SS_X$ and $SS_Y$ respectively show the standard deviation of $X$ and $Y$. Moreover, $SP_{Dxy}$ expresses the covariance between variables $X$ and $Y$. Denoted as $\rho(r)$, Pearson's correlation coefficient is a statistical evaluation spanning -1 and +1. The strength of the association between the two independent variables is found by means of these coefficient magnitudes. Furthermore, the positive and negative signs of these coefficients point accordingly to the direction of the link, either direct or reverse. While a correlation value close to 0 suggests a poor association between two independent factors, a correlation number close to 1 denotes a significant relationship between them [50]. It should be noted that the Python package Seaborn [51] was used to obtain the results presented in Fig. 2.

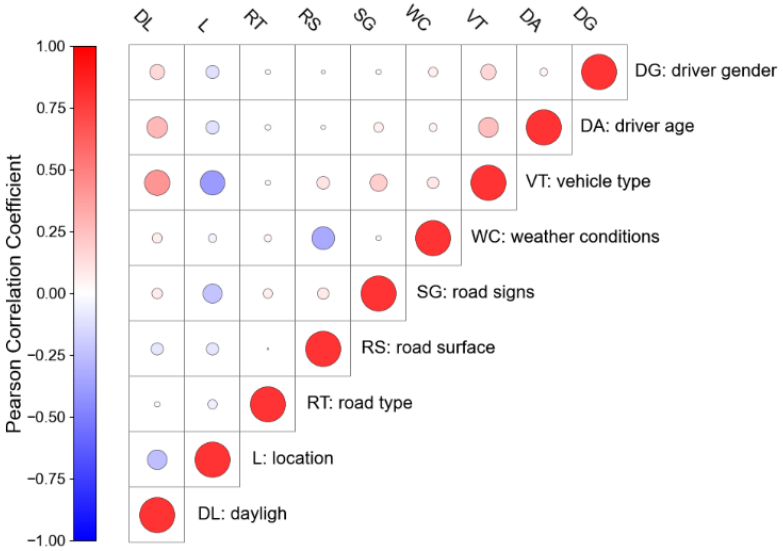



**Fig. 2** Pearson correlation matrix with circle markers

Fig. 2 shows the Pearson correlation coefficient matrix for the dataset, visualized using a grid of colored circles. The strength and nature of the link between variables including daylight (DL), location (L), road type (RT), road surface (RS), road signs (SG), weather conditions (WC), vehicle type (VT), driver age (DA), and driver gender (DG) each circle's size and color intensity reflect. Red shows positive correlations; blue shows negative correlations; the lack of a circle denotes no noteworthy correlation. We regard as "strong" any ρ whose absolute value exceeds 0.85 for modeling needs. Not one of the coefficients in this matrix surpasses this level, though. This graphic depiction highlights important relationships including the modest negative correlation between vehicle type and location ($r$ = -0.485) and the modest positive correlation between vehicle type and daylight hours ($r$ = 0.516), so enabling a quick and easy knowledge of the relationships between variables.

Without consulting numerical data, the visual signals given by the colored circles help one to recognize patterns and connections in this correlation matrix. The study of the data shown in the table suggests that most correlations are near zero, thereby implying essentially linear connections among them. For example, the rather evident moderate negative correlation between daylight and location ($r$ = -0.311) indicates that the location factor usually tends to drop as daylight rises. Correspondingly, the modest positive connection between daylight and driver age ($r$ = 0.349) suggests that older drivers are more likely to be on the road during daylight hours. For initial research, this representation is a useful tool since it helps scientists rapidly identify important relationships and concentrate on areas that might call for more study. Though two variables may be related, this does not guarantee one thing causes another.

### 5.2. SVM Modelling

This work developed a binary classification model in MATLAB using SVM to estimate the involved vehicle count. In this research, the number of accidents was the dependent variable; DL, L, RT, RS, SG, WC, VT, DA, and DG were the independent variables. The good performance of an SVM model depends on the exact control parameter selection and appropriate architecture of the model. Therefore, we aimed to find an optimum SVM model for a multi-class classification issue concerning the number of cars involved in accidents by means of an iterative hyperparameter tuning strategy, often known as manual hyperparameter tuning or a guided variation of grid search [52]. In fact, specific relationships for determining control parameters do not exist or are limited, hence; we use the try and error technique to determine them [53,54]. This procedure involved systematic adjustment of important parameters: the kernel function, the box constraint (C), and the kernel scale; i.e. for the RBF kernel. While conventional grid search investigates all hyperparameter combinations within preset ranges, our approach includes flexibility depending on insights obtained from each experimental iteration, therefore matching suggestions in the literature for successful hyperparameter tuning in SVM models [55]. Model effectiveness was found by performance throughout training, validation, and test sets. The primary goal was to optimize test set accuracy by precisely matching model complexity with generalization capacity, hence lowering the risk of overfitting [56]. The experimental path is summed in Table 2, which also offers an understanding of the justification for every hyperparameter change.

**Table 2** Empirical analysis of SVM hyperparameter tuning

| No. | Kernel function | Box constraint (C) | Kernel scale (for RBF) | Train accuracy | Validation accuracy | Test accuracy | Notes/observations |
|---|---|---|---|---|---|---|---|
| 1 | Linear | 1 (default) | - | 85.73 | 88.67 | 80.00 | Baseline linear kernel; satisfactory but suboptimal performance. |
| 2 | RBF | 1 (default) | auto | 89.20 | 84.67 | 86.00 | Transition to RBF kernel; improved test accuracy, yet evidence of mild overfitting (Train > Validation). |
| 3 | RBF | 0.1 | auto | 84.13 | 80.67 | 82.00 | Decreased C to reduce overfitting; successful in reducing overfitting but at the cost of slightly diminished test accuracy. |
| 4 | RBF | 10 | auto | 91.07 | 79.33 | 88.00 | Increased C; substantial test accuracy improvement, but overfitting exacerbated. |
| 5 | RBF | 10 | 0.1 | 92.13 | 74.00 | 72.00 | Decreased KernelScale; induced severe overfitting (high Train, low Validation/Test). |
| 6 | RBF | 10 | 10 | 86.00 | 87.33 | 84.00 | Increased KernelScale; successful overfitting mitigation, yet a minor decline in test accuracy. |
| 7 | RBF | 5 | 10 | 85.47 | 80.67 | 84.00 | Further C reduction; overfitting further curtailed, but no test accuracy gain. |
| 8 | Polynomial | 1 | N/A | 85.47 | 80.67 | 76.00 | Explored polynomial kernel; inferior test accuracy compared to RBF. |
| 9 | Polynomial | 0.1 | N/A | 90.53 | 80.67 | 82.00 | Decreased C for polynomial kernel; marginal improvement but still lags behind RBF. |
| 10 | Polynomial | 10 | N/A | 91.60 | 83.33 | 78.00 | Increased C for polynomial kernel; led to overfitting and decreased test accuracy. |
| 11 | RBF | 7.5 | 10 | 84.93 | 83.33 | 80.00 | Further C fine-tuning for RBF; overfitting reduction, but test accuracy slightly lower. |
| 12 | RBF | 8.75 | 10 | 85.47 | 83.33 | 74.00 | Additional C fine-tuning; test accuracy dropped. |
| 13 | RBF | 10 | 5 | 87.60 | 84.67 | 84.00 | Adjusted KernelScale; notable validation accuracy improvement, test accuracy maintained. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 14 | RBF | 12 | 5 | 86.80 | 86.00 | 84.00 | Slight C increase; minor Train/Validation improvement, and no change in Test. |
| 15 | RBF | 15 | 5 | 87.60 | 82.00 | 90.00 | Further increased C; significant improvement in Test accuracy, but a slight decrease in Validation. |
| 16 | RBF | 12 | 5 | 86.53 | 82.00 | 88.00 | Decreased C slightly; reduced overfitting, maintained high Test accuracy. |
| 17 | RBF | 13.5 | 5 | 86.46 | 87.33 | 86.20 | Further C fine-tuning; improved Validation, slight Test accuracy decrease. |

With test accuracy of 86.2% and a solid balance between training and validation accuracy, Experiment 17 turned out to be the most fascinating configuration based on the series of tests. This underscores a key characteristic of predictive models: their capacity to generalize appropriately to the test set [57]. Furthermore, showing less overfitting than other configurations with the same test accuracy, it suggested a more stable model. Analyzing the final model's confusion matrix and receiver operating characteristic (ROC) curve would assist us to understand its performance even further as seen in Figs. 3 and 4.

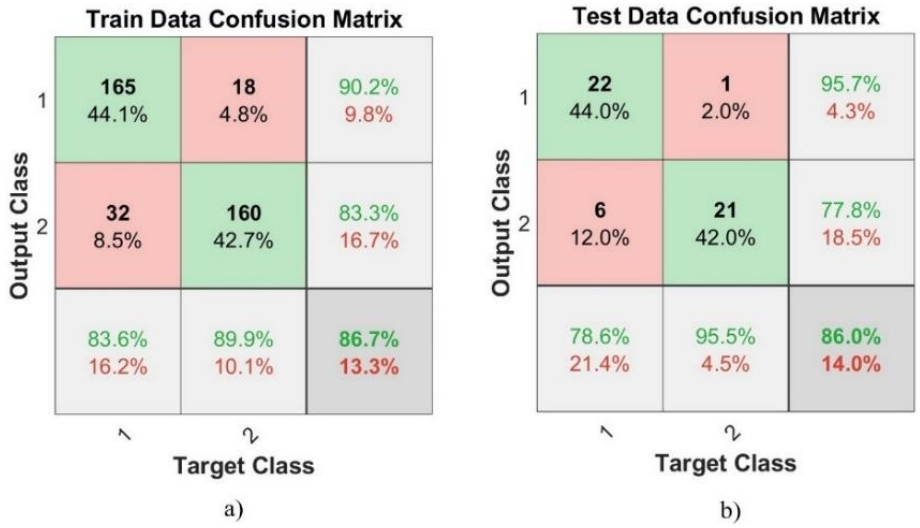



**Fig. 3** Confusion matrices for the training data (a) and testing data (b)

According to Fig. 3a, the 17th binary classification model for the train properly identified 165 occurrences of the first class labeled "1," hence implying the participation of only one vehicle in the accident. But it wrongly classified eighteen second-class cases (that is, involving two or more vehicles) as belonging to the first class designated "1".

Notably, the model was able to classify the train set with a great degree of accuracy, more especially 86%. The binary classification findings for the test data point to 21 instances belonging to the second class with the label "2" and 22 cases as properly classified as belonging to the first class with the label "1" (Fig. 3b). Under labels "1" and "2," the 17th model, however, incorrectly projected 6 instances from the second class and 1 case from the first class. In test data categorization, the 17th model therefore obtained a reasonable accuracy of 86%. SVM is shown in the paper to be a consistent method for system modeling – more notably, for accident prediction.

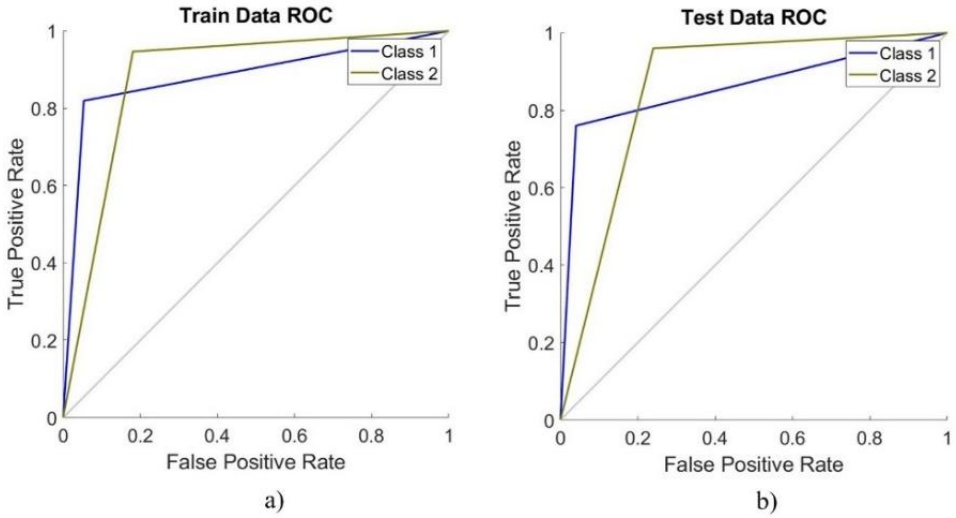



**Fig. 4** ROC curve of best SVM model for training dataset (a), and testing dataset (b)

The performance of model 17th was evaluated using the ROC curve, a probability-based curve. The model's AUC was higher than other developed models, ranging from 0 to 1. AUC values between 0 and 1 indicate the model's performance. AUC values less than 0.5 indicate the model's inadequacy, while values above 0.5 indicate the model's effectiveness in training, testing, and total ROC curve [58]. This indicates the model's superiority in classification problems.

By means of this empirical investigation, we methodically traversed the hyperparameter space employing a directed network search strategy akin to that applied in other studies on SVM-based traffic safety assessment models [59]. This resulted in the discovery of an ideal SVM configuration balancing generalization with model complexity. Based on the chosen configuration, which underlying the RBF kernel, promising performance on the test set indicates its possible efficiency in estimating the number of vehicles involved in crashes depending on supplied criteria.

### 5.3. SVM-WOA

A predictive model was developed in the MATLAB environment using a mix of WOA and SVM. The WOA algorithm was employed to optimize some parameters of the SVM in order to maximize the performance of the SVM model. In the modeling process, a random selection is made where 75% of the dataset (375 data points) is designated as the training dataset, while the remaining 25% (125 data points) is allocated as the testing dataset [60].

When using the SVM algorithm for modeling, all data with labels "1" and "2" were classified into two classes. Defining the control parameters of WOA will help one to efficiently establish and maximize the SVM model's parameters. The capacity of the model to converge fast and precisely is much influenced by these values. Although exact correlations to define these criteria are not known, past studies and professional opinions have helped to create ranges for every one of them. This covers the count of whale populations (between 5 and 60) and the iterations (between 10 and 200). The most appropriate choices then were selected by means of experimentation and improvement [61].

The model also underwent k-fold cross-validation, a technique whereby the data was split into K subgroups. One of the K-1 samples was set aside for validation at each iteration under this strategy; the other K-1 sample was set aside for training. The method was repeated K times, once for training and once for validation using every subset of data. Thus, choosing the average result of this K validation helped to define the last estimate. The determination of the k-fold value depends on the volume of data and the opinion of experts rather than on a specific technique. This present work used a k-fold value of 3. Furthermore, chosen were three distinct kernel functions: notably RBF, POL, and LIN. For SVM-WOA, a total of 48 models were constructed; 16 models were generated for each kernel function dependent on the number of control parameters. Table 3 shows 16 models related to RBF as an alternate category as the most optimal models were linked to the RBF kernel function.

**Table 3** The performance of training and testing models is determined by the different control parameters

| No. | iterations | populations | Training Accuracy (%) | Testing Accuracy (%) |
|---|---|---|---|---|
| 1 | 5 | 10 | 86.67 | 85.60 |
| 2 | 5 | 50 | 88.00 | 84.00 |
| 3 | 5 | 100 | 87.73 | 87.20 |
| 4 | 5 | 200 | 87.47 | 84.80 |
| 5 | 20 | 10 | 86.13 | 83.20 |
| 6 | 20 | 50 | 88.00 | 86.40 |
| 7 | 20 | 100 | 86.40 | 85.60 |
| 8 | 20 | 200 | 89.07 | 84.80 |
| 9 | 40 | 10 | 86.67 | 84.00 |
| 10 | 40 | 50 | 86.13 | 84.20 |
| 11 | 40 | 100 | 85.60 | 84.00 |
| 12 | 40 | 200 | 88.53 | 86.40 |
| 13 | 60 | 10 | 86.93 | 86.20 |
| 14 | 60 | 50 | 87.20 | 84.60 |
| 15 | 60 | 100 | 89.33 | 86.40 |
| 16 | 60 | 200 | 89.07 | 88.00 |

After building several models and assessing their accuracy, all the models were ranked simply using a basic ranking system [62]. The scores are given based on the accuracy percentage as a performance indicator, in a scale from 1 to 16. Here, 16 denotes the highest and 1 represents the lowest performance. These scores will be given according to the performance of each model with respect to other models. After assigning the scores for all performance metrics, the sum of the scores across all indicators gives the total score of each model. Then, ranking is done by comparing total scores to determine which model has the highest cumulative score and hence is the best performing. This will ensure that with this scoring and ranking system, multiple dimensions of performance will be considered to properly evaluate all models comprehensively and make a balanced comparison, prioritizing both training and testing accuracy. Table 4 lists the ranking results.

**Table 4** Ranking of developed models

| MLP-WOA model number | Iterations | Populations | Rating for accuracy of training | Rating for accuracy of testing | Total rank |
|---|---|---|---|---|---|
| 1 | 5 | 10 | 8 | 12 | 20 |
| 2 | 5 | 50 | 13 | 8 | 21 |
| 3 | 5 | 100 | 12 | 15 | 27 |
| 4 | 5 | 200 | 11 | 11 | 22 |
| 5 | 20 | 10 | 6 | 7 | 13 |
| 6 | 20 | 50 | 13 | 14 | 27 |
| 7 | 20 | 100 | 7 | 12 | 19 |
| 8 | 20 | 200 | 15 | 11 | 26 |
| 9 | 40 | 10 | 8 | 8 | 16 |
| 10 | 40 | 50 | 6 | 9 | 15 |
| 11 | 40 | 100 | 5 | 8 | 13 |
| 12 | 40 | 200 | 14 | 14 | 28 |
| 13 | 60 | 10 | 9 | 13 | 22 |
| 14 | 60 | 50 | 10 | 10 | 20 |
| 15 | 60 | 100 | 16 | 14 | 30 |
| 16 | 60 | 200 | 15 | 16 | 31 |

Table 4 shows that the best performance of model number 16 has a rank of 31 out of the total 15 training and 16 testing accuracy ratings. Therefore, it implies that at rank 31, the model reaches an ideal point for balancing between training and testing accuracy. The total rank of model number 15 comes next with a rank of 30 and performs less compared to model 16. The remaining models have shown various degrees of performances, the least successful mix of training and testing accuracy ratings of models with numbers 5 and 11. With the total ranking of 28, model number 12 has the third highest total ranking, indicating that it has a good balance between its training and testing accuracy. Model 16 has the best model in terms of a total rank measure for both training and testing accuracy.

Based on Fig. 5, the 16th binary classification model accurately identified 163 instances of the first class labeled "1" in an accident, indicating the involvement of only one car. However, it incorrectly classified 8 instances of the second class, indicating involvement of two or more vehicles, as "1". The model achieved a high level of accuracy, specifically

89.1%, in categorizing the training dataset. The test data showed that 52 cases were correctly identified as "1" and 58 as "2", but the model incorrectly predicted 3 cases and 12 cases from the second class. The 16th model achieved a satisfactory accuracy of 88% in classifying the test data, demonstrating the reliability of the combination of WOA and SVM for system modeling.

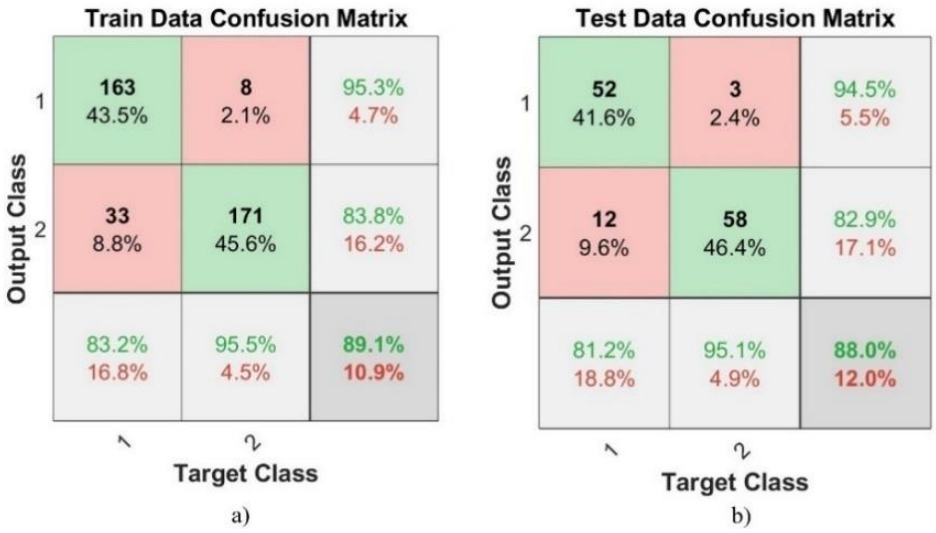



**Fig. 5** Confusion matrices for the training data (a) and testing data (b)

### 5.4. SVM-CSA

A predictive model was developed by combining CSA and SVM in the MATLAB environment. In the CSA algorithm, the optimum performance of the SVM model was achieved after optimizing its parameters. The same datasets used in the SVM analysis were used to develop SVM-CSA modeling, comprising 75% of the dataset containing 375 data as training datasets and 25% as 125 data as a test dataset [59]. All the data were labeled as two classes, which were "1" and "2". The best setting of the SVM model had been chosen in the SVM analysis section. The determination of control parameters of CSA was done, which plays an important role in the convergence of the model. Since there is no relation that precisely determines such parameters, the approach to be followed was the same as in the model SVM-WOA. Therefore, the same population number and number of iterations were considered. The best parameters were selected based on a trial-and-error approach.

In addition, the three different types of kernel functions including RBF, POL, and LIN were used. According to the number of control parameters, a total of 16 models were built and their outcomes are displayed in Table 5.

Following the construction of several models and the evaluation of their accuracy, a straightforward ranking method was employed to rank all the models [62]. The ranking outcomes are presented in Table 6.

**Table 5** The performance of training and testing models is determined by the different control parameters

| No. | iterations | populations | Training Accuracy (%) | Testing Accuracy (%) |
|---|---|---|---|---|
| 1 | 5 | 10 | 85.07 | 83.20 |
| 2 | 5 | 50 | 87.47 | 85.60 |
| 3 | 5 | 100 | 87.07 | 86.40 |
| 4 | 5 | 200 | 87.73 | 87.20 |
| 5 | 20 | 10 | 85.60 | 84.80 |
| 6 | 20 | 50 | 90.10 | 88.00 |
| 7 | 20 | 100 | 88.80 | 86.40 |
| 8 | 20 | 200 | 88.80 | 88.80 |
| 9 | 40 | 10 | 90.13 | 87.20 |
| 10 | 40 | 50 | 89.60 | 87.30 |
| 11 | 40 | 100 | 87.70 | 86.40 |
| 12 | 40 | 200 | 88.53 | 85.60 |
| 13 | 60 | 10 | 88.27 | 88.80 |
| 14 | 60 | 50 | 88.53 | 88.00 |
| 15 | 60 | 100 | 86.93 | 84.80 |
| 16 | 60 | 200 | 87.07 | 85.00 |

**Table 6** Ranking of developed models

| MLP-WOA model number | Iterations | Populations | Rating for accuracy of training | Rating for accuracy of testing | Total rank |
|---|---|---|---|---|---|
| 1 | 5 | 10 | 4 | 8 | 12 |
| 2 | 5 | 50 | 8 | 11 | 19 |
| 3 | 5 | 100 | 7 | 12 | 19 |
| 4 | 5 | 200 | 10 | 13 | 23 |
| 5 | 20 | 10 | 5 | 9 | 14 |
| 6 | 20 | 50 | 15 | 15 | 30 |
| 7 | 20 | 100 | 13 | 12 | 25 |
| 8 | 20 | 200 | 13 | 16 | 29 |
| 9 | 40 | 10 | 16 | 13 | 29 |
| 10 | 40 | 50 | 14 | 14 | 28 |
| 11 | 40 | 100 | 9 | 12 | 21 |
| 12 | 40 | 200 | 12 | 11 | 23 |
| 13 | 60 | 10 | 11 | 16 | 27 |
| 14 | 60 | 50 | 12 | 15 | 27 |
| 15 | 60 | 100 | 6 | 9 | 15 |
| 16 | 60 | 200 | 7 | 10 | 17 |

The ranking of the created models based on their performance in training and testing is presented in Table 6. The model with the highest performance is model number 6, which has a total rank of 30. This rank is determined by combining a training accuracy rating of 15 and a testing accuracy rating of 15. The model's exceptional total rank indicates that it has achieved the most optimal balance between training and testing accuracy compared to all the analyzed models. Models 8 and 9, which come after model number 6, have a total

rank of 29, making it the second highest. This model exhibits robust performance in both domains but is slightly lower than model number 6. The remaining models exhibit a variety of performance levels. Model number 1 has the lowest total rank of 12, indicating the least effective combination of training and testing accuracy ratings. Conversely, model number 10 possesses a total rank of 28, which is the third highest. This suggests that the model achieves a commendable equilibrium between training and testing accuracy.

Model number 6 is deemed the superior model based on the total rank measure, which considers both training and testing accuracy. The remaining models exhibit different levels of achievement, with certain models demonstrating exceptional training accuracy, others showcasing superior testing accuracy, and some striking a harmonious equilibrium between the two.

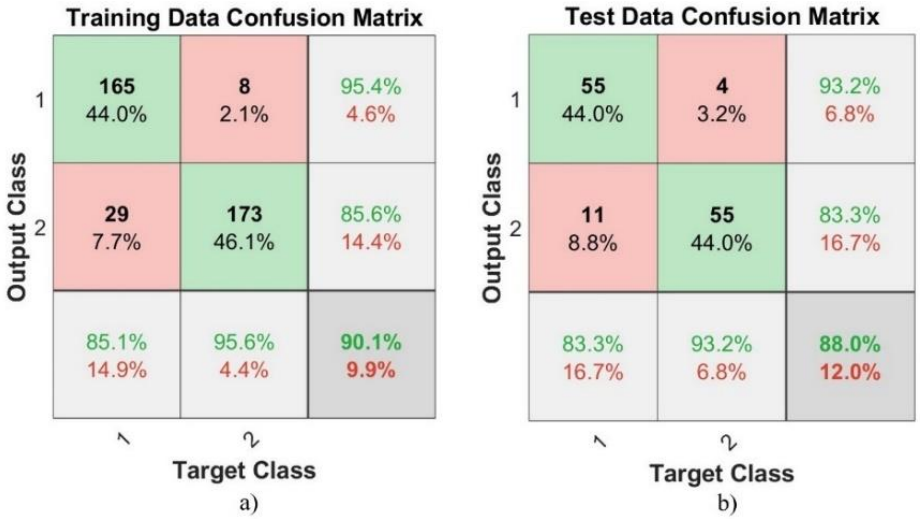



**Fig. 6** Confusion matrices for the training data (a) and testing data (b)

Based on Fig. 6a, the sixth binary classification model accurately identified 165 instances of the first class labeled "1" (indicating the involvement of only one car in the accident). However, it incorrectly classified 8 instances of the second class (indicating involvement of two or more vehicles) as belonging to the first class labeled "1". It is important to note that the model was able to categorize the training dataset with a high level of accuracy, specifically 90.1%.

According to Fig. 6b, the binary classification results for the test data indicate that 55 cases belonging to the first class with label "1" and 55 cases belonging to the second class with label "2" were correctly identified. However, the sixth model incorrectly predicted 4 cases from the first class and 11 cases from the second class, with labels "1" and "2" respectively. As a result, the sixth model achieved a satisfactory accuracy of 88% in classifying the test data. The investigation demonstrates that the combination of CSA and SVM is a dependable methodology for system modeling, specifically in anticipating accident occurrences.

### 5.5. SVM-MVO

The present work designed a prediction model using the MATLAB environment, which integrates MVO with SVM. Parameters of the SVM model were tuned using the MVO technique in order to achieve the highest performance. The SVM-MVO modeling technique utilizes the same dataset as SVM analysis, with 75% of the data allocated for training and the remaining 25% for testing [59]. Based on the entire dataset, the two categories were designated as "1" and "2". Applying the same approach as the SVM-WOA model, the control parameters of the MVO were determined through a process of rigorous experimentation. Three unique kernel functions were utilized, and a total of sixteen models were produced, depending on the number of control parameters being considered. The results of these models are presented in Table 7.

**Table 7** The performance of training and testing models is determined by the different control parameters

| No. | iterations | populations | Training Accuracy (%) | Testing Accuracy (%) |
|---|---|---|---|---|
| 1 | 5 | 10 | 85.87 | 84.00 |
| 2 | 5 | 50 | 88.53 | 86.40 |
| 3 | 5 | 100 | 88.80 | 85.60 |
| 4 | 5 | 200 | 86.13 | 84.80 |
| 5 | 20 | 10 | 87.20 | 84.80 |
| 6 | 20 | 50 | 87.73 | 86.40 |
| 7 | 20 | 100 | 88.80 | 87.20 |
| 8 | 20 | 200 | 89.60 | 87.20 |
| 9 | 40 | 10 | 85.87 | 84.00 |
| 10 | 40 | 50 | 88.27 | 87.20 |
| 11 | 40 | 100 | 88.27 | 86.40 |
| 12 | 40 | 200 | 88.00 | 87.20 |
| 13 | 60 | 10 | 86.93 | 84.00 |
| 14 | 60 | 50 | 88.27 | 87.20 |
| 15 | 60 | 100 | 88.80 | 87.00 |
| 16 | 60 | 200 | 89.00 | 87.20 |

Following the construction of several models and the evaluation of their accuracy, a straightforward ranking method was employed to rank all the models. The ranking outcomes are presented in Table 8. It displays the models' relative rankings according to their performance in both training and testing. Model 8 attained the highest level of equilibrium between training and test accuracy, resulting in a total rank of 32. Model 16, with a rating of 31, demonstrates robust performance in both domains, but slightly below that of Model 8. Model 8, widely regarded as the leading model, has distinct degrees of achievement.

According to Fig. 7, which shows the confusion matrix, the 8th binary classification model accurately identified 162 instances of the first class labeled "1" in an accident, with a high level of accuracy of 89.6%. However, it incorrectly classified 6 instances of the second class, indicating the involvement of two or more vehicles, as "1". The model was able to categorize the training dataset with a satisfactory accuracy of 87.2%. The third model achieved a satisfactory accuracy of 87.2% in classifying the test data.

**Table 8** Ranking of developed models

| MLP-WOA model number | Iterations | Populations | Rating for accuracy of training | Rating for accuracy of testing | Total rank |
|---|---|---|---|---|---|
| 1 | 5 | 10 | 6 | 11 | 17 |
| 2 | 5 | 50 | 13 | 14 | 27 |
| 3 | 5 | 100 | 14 | 13 | 27 |
| 4 | 5 | 200 | 7 | 12 | 19 |
| 5 | 20 | 10 | 9 | 12 | 21 |
| 6 | 20 | 50 | 10 | 14 | 24 |
| 7 | 20 | 100 | 14 | 16 | 30 |
| 8 | 20 | 200 | 16 | 16 | 32 |
| 9 | 40 | 10 | 6 | 11 | 17 |
| 10 | 40 | 50 | 12 | 16 | 28 |
| 11 | 40 | 100 | 12 | 14 | 26 |
| 12 | 40 | 200 | 11 | 16 | 27 |
| 13 | 60 | 10 | 8 | 11 | 19 |
| 14 | 60 | 50 | 12 | 16 | 28 |
| 15 | 60 | 100 | 14 | 15 | 29 |
| 16 | 60 | 200 | 15 | 16 | 31 |

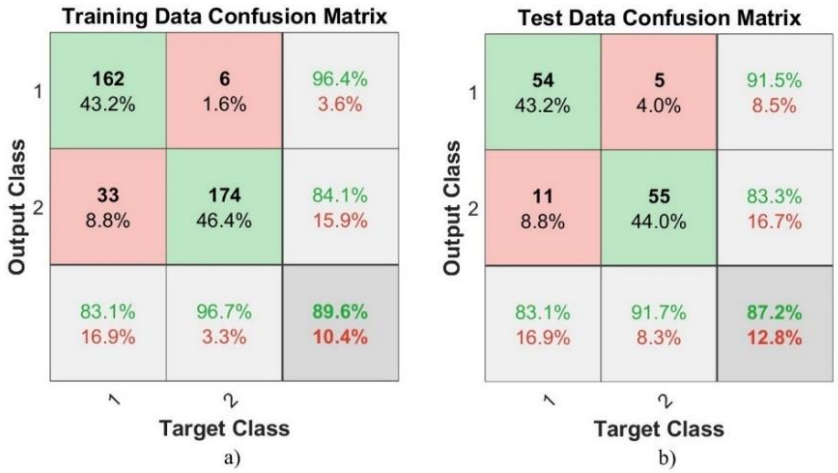
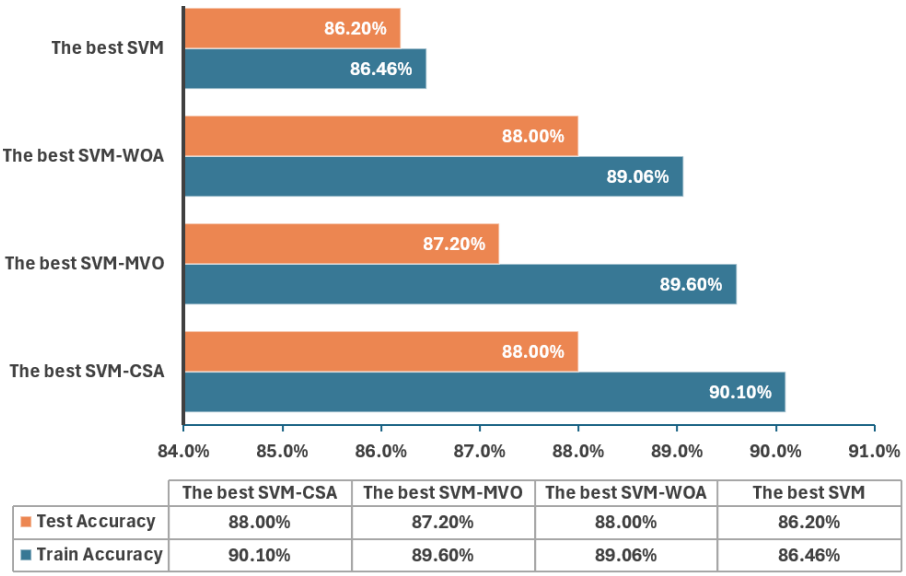



**Fig. 7** Confusion matrices for the training data (a) and testing data (b)

## 6. RESULT AND DISCUSSION

Each year, many people die as a result of road traffic crashes. Therefore, knowing the impact of various contributing factors on the number of road accidents and taking the necessary measures to reduce accidents can have a significant impact on increasing the level of road safety. In this research, four ML methods, namely SVM, SVM-WOA, SVM-CSA, and SVM-MVO were employed to conduct the binary classification modeling. After multiple modeling, the best model was selected based on the accuracy of the modeling

performance. A comparison was made between the best models based on the accuracy of training and testing, as shown in Fig. 8.




| | The best SVM-CSA | The best SVM-MVO | The best SVM-WOA | The best SVM |
|---|---|---|---|---|
| Test Accuracy | 88.00% | 87.20% | 88.00% | 86.20% |
| Train Accuracy | 90.10% | 89.60% | 89.06% | 86.46% |

**Fig. 8** Comparison between the best SVM, SVM-CSA, MLP-WOA, and SVM-MVO models for training and testing accuracies

Based on Fig. 8, the SVM-CSA model demonstrated superior performance compared to the other best models in forecasting the number of accidents. It achieved training and testing accuracies of 90.1% and 88%, respectively, which were greater than the accuracies of the other three models. It is important to note that all models demonstrated satisfactory levels of accuracy and robustness. Therefore, it can be inferred that they are dependable modeling systems for predicting the occurrence of crashes. These models can serve as valuable tools for analyzing road safety in the field of transportation engineering.

The occurrence of road accidents can result in significant economic and human detriment to society. Consequently, evaluating the influence of factors that affect the frequency of accidents can offer engineers engaged in road safety management a comprehensive understanding. To evaluate the influence of daylight (DL), location (L), road type (RT), road surface (RS), road signs (SG), weather conditions (WC), vehicle type (VT), driver age (DA), and driver gender (DG) on estimating the number of cars involved in accidents, a sensitivity analysis was conducted. This sensitivity analysis utilizes the cosine amplitude approach as described by [63]:

$$r_{ij} = \frac{\sum_{k=1}^{n}(x_{ik} \times y_{jk})}{\sqrt{\sum_{k=1}^{n} x_{ik}^2 \sum_{k=1}^{n} y_{ik}^2}} \tag{33}$$

where $r_{ij}$ represents the strength of the relationship, $n$ denotes the number of datasets, and $x_{ik}$ and $y_{ij}$ represent the input parameters and anticipated output, respectively.

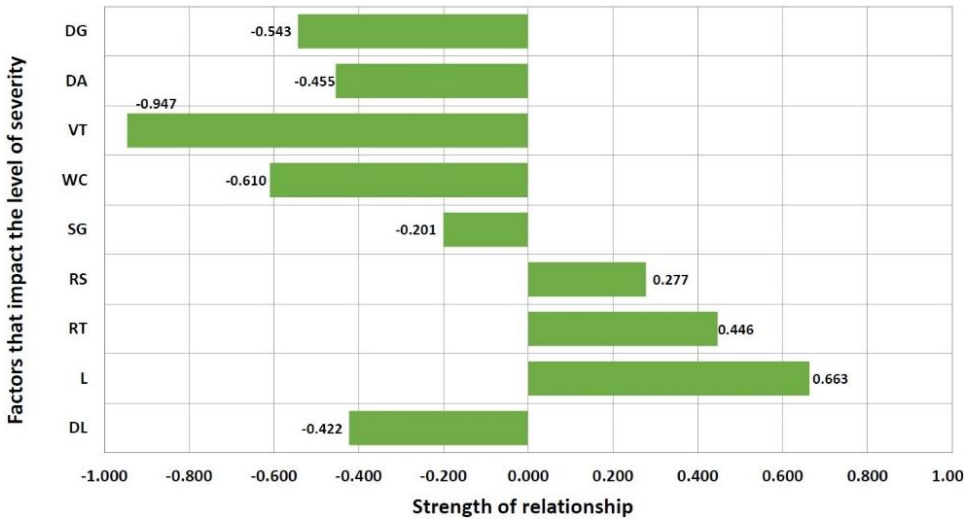



**Fig. 9** Results of sensitivity analysis obtained from the SVM-CSA model

Additionally, the individual conditional expectation (ICE) plot (Fig. 10) was used to achieve a deeper understanding of the relationship between the most influential predictor variable (vehicle type) identified in Fig. 9 and the expected result (number of vehicles involved in an accident). ICE plots show, for individual instances, how a model prediction would change if the value of one feature varied, while the other features remained constant [64].

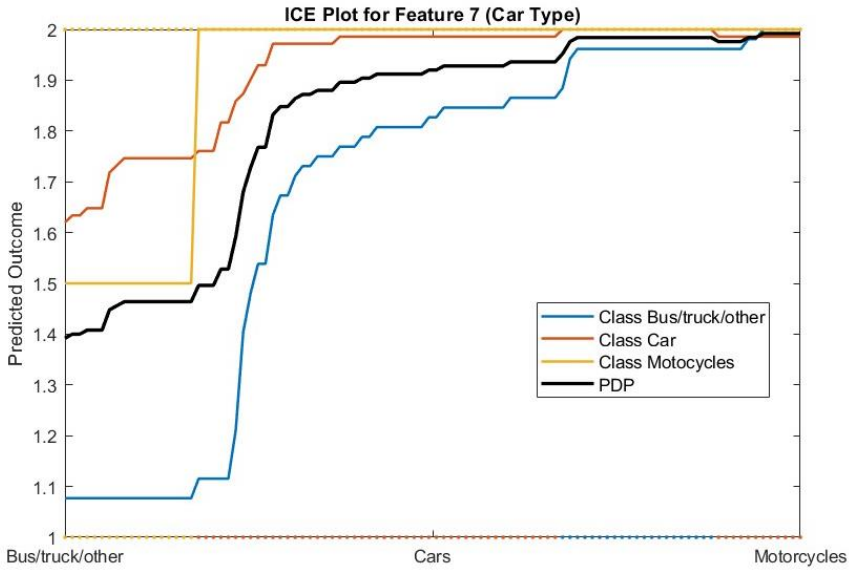



**Fig. 10** Individual conditional expectation plot for vehicle type and multi-vehicle accident probability

The ICE plot visualizes the effect of vehicle type on the probability of multi-vehicle accidents. At an aggregated level, the increase in probability tends to follow the change from "Bus/truck/other" via "Cars" towards "Motorcycles." However, the individual ICE lines behave quite differently, showing that this is moderated by other factors that vary across the instances. In the ICE plot, the three-colored lines correspond with the three categories of vehicle type: blue for "Bus/truck/other", red for "Cars", and orange for "Motorcycles". Also, the Partial Dependence Plot (PDP) (black line) shows the average effect of one feature on model predictions, holding all other features constant. The PDP is an overall picture of the dependence that the model has on a particular feature and is obtained by averaging the lines from ICE plots [64, 65]. Figs. 9 and 10 demonstrate that the model analysis yields results reflecting the influence of the examined components, hence confirming the dependability of the findings. Furthermore, the following remarks can also be made.

The vehicle type was found to have the most imperative impact on the number of cars involved in accidents. This points to the indication that some types of vehicles are more prone to accidents and could further enhance the devastating effects of these accidents as well [61]. Indeed, further confirmation of this finding has been made through ICE plot analysis, which shows that the highest average predicted probability of involvement in multi-vehicle accidents is for motorcycles, followed by cars, and then buses/trucks/other vehicles. The fact that the type of vehicle has a relatively consistent effect on the prediction for motorcycles, given the diversity of instances in this class, would therefore imply inherent characteristics or patterns of traffic that put them at higher risk. This is especially present in multi-vehicle accidents. With the increase in electric motorbikes and scooters, the demand for focused interventional policies, such as customized training programs and improvements in the road structure, also increases [66-68].

Therefore, we can conclude that this would form a possible basis on which specific traffic patterns may be examined as a function of the vehicle type. Motorcycles, having the highest relative risk, might be expected to have accidents due to their lane changing and higher speeds through congested traffic. Such accidents, in particular, could be radically minimized through the introduction of advanced technologies related to adaptive cruise control and warning systems. Such an approach would result in fewer road accidents, more even flow conditions of traffic, and reduced costs arising from these road accidents.

Furthermore, location is also a very important factor; i.e. topography and features of any area or a part of the road can result in more frequent accidents within these areas, requiring further research and focused solutions. Bad weather, like rain or fog, also has a great impact on accident rates; that is the meaning of teaching drivers how to behave in traffic management systems capable of adapting to diverse situations [66, 69]. Driving factors, such as gender and age, further influence involvement in accidents; this might be due to differences in either driving style or the tendency toward risk exposure. These factors could be addressed through strategic teaching programs and sensitization processes effectively [66]. The nature of the road is another factor, with different roads being risky to different measures. Improved road design, traffic control, and signage can help drastically reduce these risks [66]. In general, diurnal daylight conditions, with increased visibility at night, accentuate the occurrence of accidents. Therefore, priorities regarding improved street lighting, marks for visibility, and awareness campaigns are needed [70].

Additionally, road surface conditions and signage, although having a relatively smaller impact compared to other factors, still play a role in road safety. Guaranteeing safety to

drivers actually depends mainly on proper maintenance, appropriate signs, and an unobstructed view [71-73].

One of the significant limitations in the proposed approach is that it cannot operate with incomplete data; therefore, full and accurate data availability is crucial for the successful implementation of this approach. Moreover, the results and interpretations obtained from this study relate only to the urban road conditions of Calabria; thus, applications in other areas or other types of road scenarios are not directly possible. Nevertheless, this research proved that classification techniques can be a strong tool in the prediction of the number of vehicles involved in traffic accidents, thus providing important information to transportation engineering.

Finally, some of the future directions of research could be extension of applicability by testing the proposed approach in a wide range of regions and road conditions to enhance generalizability. Future studies may investigate the implications of emerging trends faced by accident prediction models with respect to electric scooters and autonomous vehicles. The interaction between environmental factors, vehicle types, and driver characteristics would provide deeper insights into the causation of an accident. Longitudinally, evaluation studies that determine the long-term effectiveness of interventions such as infrastructure improvement, training programs, and publicity should help revise and prioritize road safety measures.

## 7. CONCLUSIONS

This study applied a binary classification approach to model the relationship between various factors and the number of vehicles involved in road accidents in the urban area of Cosenza, southern Italy. Variables such as DL, L, RT, RS, SG, WC, VT, DA, and DG were considered as input factors for the analysis. For this purpose, four machine learning models, namely SVM, SVM-CSA, SVM-WOA, and SVM-MVO, have been developed and compared to choose the best predictive model. The classification process distinguished between single-vehicle and multi-vehicle accidents. Model performance was assessed using confusion matrices and ROC curves. Among the evaluated models, SVM-CSA achieved the best performance, with a 90.1% increase in training accuracy and a high test accuracy of 88%, outperforming the other techniques. The sensitivity analysis revealed that the factor "vehicle type" had the highest influence, while "road signs" had the least. These findings underline the effectiveness of advanced machine learning techniques in establishing the key accident-related factors, therefore providing useful insights into improvement in road safety and transportation safety management strategies.

## REFERENCES

1. Guido, G., Haghshenas, S.S., Haghshenas, S.S., Vitale, A., Astarita, V., Haghshenas, A.S., 2020, *Feasibility of stochastic models for evaluation of potential factors for safety: a case study in Southern Italy,* Sustainability, 12(18), 7541.
2. Ristić, B., Bogdanović, V., Stević, Ž., Marinković, D., Papić, Z., Gojković, P., 2024, *Evaluation of Pedestrian Crossings Based on the Concept of Pedestrian Behavior Regarding Start-Up Time: Integrated Fuzzy MCDM Model*, Tehnički Vjesnik, 31(4), pp. 1206-1214.
3. Saafi, N., Dhouib, K., 2024, *An Ontological Model to Enhance Traffic Conditions in Smart City Domain*, Spectrum of Engineering and Management Sciences, 2(1), pp. 70-84.

4. Mondal, S., Goswami, S. S., 2024, *Rise of Intelligent Machines: Influence of Artificial Intelligence on Mechanical Engineering Innovation*, Spectrum of Engineering and Management Sciences, 2(1), pp. 46-55.
5. Korovkinas, K., Danėnas, P., Garšva, G., 2020, *Support vector machine parameter tuning based on particle swarm optimization metaheuristic*, Nonlinear Analysis: Modelling and Control, 25(2), pp. 266-281.
6. Ul Haq, H. B., Younis, R., Ali, M. S., 2025, *Towards Robust Network Security: Evaluating Machine Learning Algorithms for Intrusion Detection*, Decision Making Advances, 3(1), pp. 126-138.
7. Cong, Y., Wang, J., Li, X., 2016, *Traffic flow forecasting by a least squares support vector machine with a fruit fly optimization algorithm*, Procedia Engineering, 137, pp. 59-68.
8. Mokhtarimousavi, S., Anderson, J.C., Azizinamini, A., Hadi, M., 2019, *Improved support vector machine models for work zone crash injury severity prediction and analysis*, Transportation research record, 2673(11), pp. 680-692.
9. Elamrani Abou Elassad, Z., Mousannif, H., Al Moatassime, H., 2020, *Class-imbalanced crash prediction based on real-time traffic and weather data: A driving simulator study*, Traffic injury prevention, 21(3), pp. 201-208.
10. Baig, M. H. M., Ul Haq, H. B., Habib, W., 2024, *A Comparative Analysis of AES, RSA, and 3DES Encryption Standards based on Speed and Performance*, Management Science Advances, 1(1), pp. 20-30.
11. Sipos, T., Afework Mekonnen, A., Szabó, Z., 2021, *Spatial econometric analysis of road traffic crashes*, Sustainability, 13(5), 2492.
12. Lord, D., Mannering, F., 2010, *The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives*, Transportation research part A: policy and practice, 44(5), pp. 291-305.
13. Vedagiri, P., Killi, D.V., 2015, *Traffic safety evaluation of uncontrolled intersections using surrogate safety measures under mixed traffic conditions*, Transportation research record, 2512(1), pp. 81-89.
14. Sawalha, Z., Sayed, T., 2001, *Evaluating safety of urban arterial roadways*, Journal of Transportation Engineering, 127(2), pp. 151-158.
15. Dell'Acqua, G., Russo, F., 2011, *Safety performance functions for low-volume roads*, The Baltic Journal of Road and Bridge Engineering, 6(4), pp. 225-234.
16. Li, L., Shrestha, S., Hu, G., 2017, June, *Analysis of road traffic fatal accidents using data mining techniques*, In 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA), pp. 363-370.
17. Pourroostaei Ardakani, S., Liang, X., Mengistu, K.T., So, R.S., Wei, X., He, B., Cheshmehzangi, A., 2023, *Road car accident prediction using a machine-learning-enabled data analysis*, Sustainability, 15(7), 5939.
18. Wang, S., Zhang, J., Li, J., Miao, H., Cao, J., 2021, *Traffic accident risk prediction via multi-view multi-task spatio-temporal networks*, IEEE Transactions on Knowledge and Data Engineering, 35(12), pp. 12323-12336.
19. Guido, G., Shaffiee Haghshenas, S., Shaffiee Haghshenas, S., Vitale, A., Astarita, V., 2022, *Application of feature selection approaches for prioritizing and evaluating the potential factors for safety management in transportation systems*, Computers, 11(10), 145.
20. Deretić, N., Stanimirović, D., Awadh, M.A., Vujanović, N., Djukić, A., 2022, *SARIMA modelling approach for forecasting of traffic accidents*, Sustainability, 14(8), 4403.
21. Dimitrijevic, B., Khales, S.D., Asadi, R., Lee, J., 2022, *Short-term segment-level crash risk prediction using advanced data modeling with proactive and reactive crash data*, Applied Sciences, 12(2), 856.
22. Danesh, A., Ehsani, M., Moghadas Nejad, F., Zakeri, H., 2022, *Prediction model of crash severity in imbalanced dataset using data leveling methods and metaheuristic optimization algorithms*, International journal of crashworthiness, 27(6), pp. 1869-1882.
23. Dong, C., Xie, K., Sun, X., Lyu, M., Yue, H., 2019, *Roadway traffic crash prediction using a state-space model based support vector regression approach*, PloS one, 14(4), e0214866.
24. Mannering, F.L., Bhat, C.R., 2014, *Analytic methods in accident research: Methodological frontier and future directions*, Analytic methods in accident research, 1, pp. 1-22.
25. Sufian, M.A., Varadarajan, J., Niu, M., 2024, *Enhancing prediction and analysis of UK road traffic accident severity using AI: integration of machine learning, econometric techniques, and time series forecasting in public health research*, Heliyon, 10(7), e28547.
26. Panda, C., Mishra, A. K., Dash, A. K., Nawab, H., 2023, *Predicting and explaining severity of road accident using artificial intelligence techniques, SHAP and feature analysis.* International journal of crashworthiness, 28(2), pp. 186-201.
27. Tselentis, D. I., Papadimitriou, E., van Gelder, P., 2023, *The usefulness of artificial intelligence for safety assessment of different transport modes.* Accident Analysis and Prevention, 186, 107034.
28. Howlader, M. M., Bhaskar, A., Yasmin, S., Haque, M. M., 2024, *A bivariate, non-stationary extreme value model for estimating opposing-through crash frequency by severity by applying artificial intelligence-based video analytics.* Transportation research part C: emerging technologies, 160, 104509.

29. Mzili, T., Mzili, I., Riffi, M. E., Pamucar, D., Simic, V., Abualigah, L., Almohsen, B., 2024, *Hybrid genetic and penguin search optimization algorithm (GA-PSEOA) for efficient flow shop scheduling solutions*, Facta Universitatis-Series Mechanical Engineering, 22(1), pp. 077-100.
30. Cortes, C., Vapnik, V., 1995, *Support-vector networks*, Machine learning, 20, pp. 273-297.
31. Yan, X., Jia, M., 2018, *A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing*, Neurocomputing, 313, pp. 47-64.
32. Maldonado, S., López, J., Jimenez-Molina, A., Lira, H., 2020, *Simultaneous feature selection and heterogeneity control for SVM classification: An application to mental workload assessment*, Expert Systems with Applications, 143, 112988.
33. Zhou, J., Qiu, Y., Zhu, S., Armaghani, D.J., Li, C., Nguyen, H., Yagiz, S., 2021, *Optimization of support vector machine through the use of metaheuristic algorithms in forecasting TBM advance rate*, Engineering Applications of Artificial Intelligence, 97, 104015.
34. Zeng, J., Roussis, P.C., Mohammed, A.S., Maraveas, C., Fatemi, S.A., Armaghani, D.J., Asteris, P.G., 2021, *Prediction of peak particle velocity caused by blasting through the combinations of boosted-CHAID and SVM models with various kernels*, Applied Sciences, 11(8), 3705.
35. Jahed Armaghani, D., Asteris, P.G., Askarian, B., Hasanipanah, M., Tarinejad, R., Huynh, V.V., 2020, *Examining hybrid and single SVM models with different kernels to predict rock brittleness*, Sustainability, 12(6), 2229.
36. Zinno, R., Haghshenas, S. S., Guido, G., VItale, A., 2022, *Artificial intelligence and structural health monitoring of bridges: A review of the state-of-the-art*, IEEE Access, 10, pp. 88058-88078.
37. Hanifehnia, J., Esmaeilzadeh, A., Mikaeil, R., Atalou, S., 2024, *Prediction of blast-induced flyrock by using neural-imperialist competitive method (Case Study: Sungun Copper Mine)*, Rudarsko-geološko-naftni zbornik, 39(5), pp. 109-120.
38. Golafshani, E. M., Behnood, A., Kim, T., Ngo, T., Kashani, A., 2024, *Metaheuristic optimization based-ensemble learners for the carbonation assessment of recycled aggregate concrete*, Applied Soft Computing, 159, 111661.
39. Kandiri, A., Ghiasi, R., Nogal, M., Teixeira, R., 2024, *Travel time prediction for an intelligent transportation system based on a data-driven feature selection method considering temporal correlation*, Transportation Engineering, 18, 100272.
40. Qiu, S., Ge, H., Li, Z., Gao, Z., Ai, C., 2024, *Network-level crash risk analysis using large-scale geometry features*, Accident Analysis & Prevention, 207, 107746.
41. Mirjalili, S., Lewis, A., 2016, *The whale optimization algorithm*, Advances in engineering software, 95, pp. 51-67.
42. Haghshenas, S. S., Guido, G., Haghshenas, S. S., Astarita, V., 2025, *Predicting the level of road crash severity: A comparative analysis of logit model and machine learning models*, Transportation Engineering, 20, 100323.
43. Yang, X.S., Deb, S., 2009, December, *Cuckoo search via Lévy flights*, In 2009 World congress on nature & biologically inspired computing (NaBIC), pp. 210-214.
44. Mantegna, R.N., 1994, *Fast, accurate algorithm for numerical simulation of Levy stable stochastic processes*, Physical Review E, 49(5), 4677.
45. Yu, X., Luo, W., 2023, *Reinforcement learning-based multi-strategy cuckoo search algorithm for 3D UAV path planning*, Expert Systems with Applications, 223, 119910.
46. Mirjalili, S., Mirjalili, S.M., Hatamlou, A., 2016, *Multi-verse optimizer: a nature-inspired algorithm for global optimization*, Neural Computing and Applications, 27, pp. 495-513.
47. Guido, G., Shaffiee Haghshenas, S., Shaffiee Haghshenas, S., Vitale, A., Astarita, V., Park, Y., Geem, Z.W., 2022, *Evaluation of contributing factors affecting number of vehicles involved in crashes using machine learning techniques in rural roads of Cosenza, Italy*, Safety, 8(2), 28.
48. Fiorini Morosini, A., Shaffiee Haghshenas, S., Shaffiee Haghshenas, S., Geem, Z.W., 2020, *Development of a binary model for evaluating water distribution systems by a pressure driven analysis (PDA) approach*, Applied Sciences, 10(9), 3029.
49. Feng, X., Li, S., Yuan, C., Zeng, P., Sun, Y., 2018, *Prediction of slope stability using naive Bayes classifier*, KSCE Journal of Civil Engineering, 22, pp. 941-950.
50. Hosseini, S. M., Ataei, M., Khalokakaei, R., Mikaeil, R., Haghshenas, S. S., 2019, *Investigating the role of the cooling and lubricant fluids on the performance of cutting disks (case study: hard rocks)*, Rudarsko-geološko-naftni zbornik, 34(2), pp. 13-25.
51. Waskom, M. L., 2021, *Seaborn: statistical data visualization*, Journal of Open Source Software, 6(60), 3021.
52. Bergstra, J., Bardenet, R., Bengio, Y., Kégl, B., 2011, *Algorithms for hyper-parameter optimization*, Proceedings of the 24th International Conference on Neural Information Processing Systems, 2011, Granada, Spain.
53. Kaur, R., Roul, R. K., Batra, S., 2023, *An efficient approach for accident severity classification in smart transportation system*, Arabian Journal for Science and Engineering, 48(8), pp. 9645-9659.
54. Haghshenas, S. S., Guido, G., Vitale, A., Astarita, V., 2023, *Assessment of the level of road crash severity: comparison of intelligence studies*, Expert systems with applications, 234, 121118.

55. Cui, J., Zhang, H., Zhao, J., Zhang, Y., 2019, *Research on SVM-based highway traffic safety evaluation model*, In Green Intelligent Transportation Systems: Proceedings of the 8th International Conference on Green Intelligent Transportation Systems and Safety. pp. 799-809.
56. Alqahtani, H., Kumar, G., 2024, *Machine learning for enhancing transportation security: A comprehensive analysis of electric and flying vehicle systems*, Engineering Applications of Artificial Intelligence, 129, 107667.
57. Cawley, G.C., Talbot, N.L., 2010, *On over-fitting in model selection and subsequent selection bias in performance evaluation*, The Journal of Machine Learning Research, 11, pp. 2079-2107.
58. Guido, G., Haghshenas, S.S., Haghshenas, S.S., Vitale, A., Gallelli, V., Astarita, V., 2020, *Development of a binary classification model to assess safety in transportation systems using GMDH-type neural network algorithm*, Sustainability, 12(17), 6735.
59. Fawcett, T., 2006, *An introduction to ROC analysis*, Pattern recognition letters, 27(8), pp. 861-874.
60. Looney, C. G., 1996, *Advances in feedforward neural networks: demystifying knowledge acquiring black boxes*, IEEE Transactions on Knowledge and Data Engineering, 8(2), pp. 211-226.
61. Zhou, J., Zhu, S., Qiu, Y., Armaghani, D.J., Zhou, A., Yong, W., 2022, *Predicting tunnel squeezing using support vector machine optimized by whale optimization algorithm*, Acta Geotechnica, 17(4), pp. 1343-1366.
62. Zorlu, K., Gokceoglu, C., Ocakoglu, F., Nefeslioglu, H.A., Acikalin, S.J.E.G., 2008, *Prediction of uniaxial compressive strength of sandstones using petrography-based models*, Engineering Geology, 96(3-4), pp. 141-158.
63. Yang, Y., Zhang, Q., 1997, *A hierarchical analysis for rock engineering using artificial neural networks*, Rock mechanics and rock engineering, 30, pp. 207-222.
64. Talebi, E., Rogers, W.P., Morgan, T., Drews, F.A., 2021, *Modeling mine workforce fatigue: Finding leading indicators of fatigue in operational data sets*, Minerals, 11(6), 621.
65. Molnar, C., 2020, *Interpretable machine learning*, Lean Publishing, Victoria, British Columbia, Canada.
66. Lee, D., Guldmann, J.M., von Rabenau, B., 2023, *Impact of driver's age and gender, built environment, and road conditions on crash severity: a logit modeling approach*, International journal of environmental research and public health, 20(3), 2338.
67. Jimenez, A., Bocarejo, J.P., Zarama, R., Yerpez, J., 2015, *A case study analysis to examine motorcycle crashes in Bogota, Colombia*, Journal of safety research, 52, pp. 29-38.
68. Stigson, H., Malakuti, I., Klingegård, M., 2021, *Electric scooters accidents: Analyses of two Swedish accident data sets*, Accident Analysis & Prevention, 163, 106466.
69. Faria, M.V., Baptista, P.C., Farias, T.L., Pereira, J.M., 2020, *Assessing the impacts of driving environment on driving behavior patterns*, Transportation, 47(3), pp. 1311-1337.
70. Carey, R.N., Sarma, K.M., 2017, *Impact of daylight saving time on road traffic collision risk: a systematic review*, BMJ open, 7(6), e014319.
71. Pérez-Fortes, A.P., Giudici, H., 2022, *A recent overview of the effect of road surface properties on road safety, environment, and how to monitor them*, Environmental science and pollution research, 29(44), pp. 65993-66009.
72. Babić, D., Fiolić, M., Babić, D., Gates, T., 2020, *Road markings and their impact on driver behaviour and road safety: A systematic review of current findings*, Journal of advanced transportation, 2020(1), 7843743.
73. Babić, D., Babić, D., Fiolic, M., Ferko, M., 2022, *Road markings and signs in road safety*, Encyclopedia, 2(4), pp. 1738-1752.