Journal of Scientific & Industrial Research
Vol. 80, November 2021, pp. 1001-1007

Mortality Prediction of Victims in Road Traffic Accidents (RTAs) in India using Opposite Population SGO-DE based Prediction Model

Junali Jasmine Jena and Suresh Chandra Satapathy*
School of Computer Engineering, KIIT (Deemed to be University), Bhubaneswar, India 751 024

Received 08 September 2021; revised 06 October 2021; accepted 21 October 2021

Getting immediate and appropriate care for the victims of Road Traffic Accidents (RTAs) in countries like India with huge population is a challenging job. In this paper a new hybridized evolutionary algorithm has been proposed for hyper-parameter tuning of the hyper-parameters of the prediction models using which mortality prediction of victims of RTAs in India have been performed. The proposed methodology Opp-SGO-DE has been used for parameter tuning in prediction algorithms like Random Forest (RF) and Support Vector Machine (SVM) and promising results were found from the experimentation. In RF, accuracy was increased from 0.75 to 0.82 and F1-score was increased from 0.66 to 0.77 in dataset-1 and accuracy was increased from 0.66 to 0.75 and F1-score was increased from 0.62 to 0.65 in dataset-2. In SVM, accuracy was increased from 0.63 to 0.74 and F1-score was increased from 0.58 to 0.67 in dataset-1 and accuracy was increased from 0.56 to 0.62 and F1-score was increased from 0.54 to 0.575 in dataset-2.

Keywords: Opp-SGO-DE, Parameter Tuning, Random Forest, Support Vector Machine

Introduction

Road Traffic Accidents (RTAs) are listed as one of the major concerns in public health issues worldwide. The global status report on road safety 2018 by the World Health Organization[1] reported the RTA as the eighth leading cause of deaths globally and thus several policies and laws have been formulated and reformed time to time for safety of individuals while driving as a precautionary measure to RTAs. Undoubtedly this is essential but the necessity of the standardized measures needed to be taken, after happening of a RTA, cannot be overlooked because from the analysis it was found that most of the deaths were preventable if patient would have received proper treatment on time. In developing and over populated countries like India, providing proper medical facilities to all the masses is a big challenge. In case of RTAs, diagnosis happen after the victim is taken to the hospital and many a times, victim has to shift to other hospitals due to lack of certain advanced medical facilities. So, in case of fatal injuries, many victims lose their life in between this transition time from the accident spot to hospital or moving in between the hospitals. There are no significant provisions for providing immediate medical attention to the fatal patients and it is obvious too, as doctors cannot be there with each ambulance. The deaths caused in RTAs is more in cities than the rural areas as cities have more traffic but the problem faced in rural areas is the non-availability of immediate medical facility as hospitals are not nearby. Thus irrespective of the locations, availing adequate and immediate treatment to the victims of RTAs is a major challenge. With high population growth rates, increasing mobility, and growing numbers of vehicles, tremendous change has occurred to the road transportation network in India over the years and hence significant rate of growth in deaths due to RTAs could be observed. Technical intervention in this field is definitely a major demand in this time.

Many researchers have worked on finding potential solution to this problem. Boo & Choi[2] in their paper showed a comparative study of four prediction models (Logistic Regression, Random Forest, Linear SVM and RBF-SVM) for mortality determination in RTAs. Similarly, Alharbi *et al.*[3] reviewed various factors for predicting mortality in traumatic injured patients of RTAs. Samad *et al.*[4] proposed a technique to predict injury severity score in relation to morbidity and mortality in RTAs. Kenneth *et al.*[5] statistically analyzed regression techniques for modelling of RTAs in Nigeria. Similarly, Vipin & Rahul[6] analyzed mortality in RTAs based on time of occurrence in Kerala.

From the literature it can be witnessed that prediction models play a major role in analyzing and

*Author for Correspondence
E-mail: suresh.satapathyfcs@kiit.ac.in

predicting important factors in RTAs. So accuracy of prediction models highly affect the efficacy of the work. Evolutionary algorithms have been largely used for parameter tuning of prediction models to increase their accuracy. Social Group Optimization (SGO) is one of such algorithm proposed by Satapathy & Naik[7] which is based on social behavior of human beings. It has been successfully applied to several real world problems such as solving multi-objective problems8, resource scheduling in cloud9, segmentation of skin melanoma images10, segmentation of brain MRI images11, short-term hydrothermal scheduling[12], brain tumor evaluation tool[13], automated detection of covid-19 infection[14], Antenna array synthesis[15], Transformer fault analysis[16], engineering design problems[17], solving Travelling Salesman problem[18], structural health monitoring in civil engineering[19] and many more.

In this paper, SGO7 has been hybridized with Differential Evolution (DE) proposed by Karaboga & Okdem[20] along with the concept of Opposite population (Opp-SGO-DE) for parameter tuning of prediction models such as Support Vector Machine (SVM) and Random Forest (RF) for predicting the mortality and mortality analysis in RTAs victims and results found were promising. No such work of application of prediction algorithms on the used datasets could be found in the literature which proves the novelty and usefulness of the work.

Proposed Methodology

Social Group Optimization[7] (SGO) is an evolutionary technique where a candidate solution represents various behavioral traits of a person and how they interact and improve themselves in the society. It is popular due to its convergence capability. For more details regarding SGO, original article may be referred. But the major limitation of SGO is more number of fitness evaluations in one run of the algorithm. When evolutionary algorithms are used for parameter tuning of the prediction models, generally, the accuracy from the prediction models serves as one of the parameter of fitness function. So, for every fitness evaluation, prediction algorithm runs completely and returns the accuracy. More number of fitness evaluation means more number of calls to prediction algorithm and more increase in total time complexity. To overcome this factor, SGO has been hybridized with Differential Evolution[20] (DE) by which the numbers of fitness evaluations are reduced to half. For more details regarding DE, the original paper may be referred.

Parameters of prediction models may have different range values and to deal with such situations normalization is done in general, but in the proposed methodology instead of normalization, shuffling of values has been done using binomial distribution policy of DE. Moreover to increase the exploration capability, concept of opposite population has been used to generate new values. Opposite population is generated using opposite number, which is a technique to generate an opposite value in the same n-dimensional space as the original value. For better understanding on generation of Opposite population, the work by Rahnamayan *et al*.[21] may be referred.

Opp-SGO-DE Algorithm

Let A_i be a candidate solution of the randomly initialized population where $i = 1, 2\ldots n$ and *'n'* is the value of total population and *'j'* is the dimension of every candidate solution where $j = 1, 2\ldots m$ and *'m'* is the total dimension or total number of variables. Every variable can have different range of values. Range of j^{th} variable is denoted by $[r_{min(j)}, r_{max(j)}]$. *'G_{best}'* is the candidate solution having best fitness value. In this paper, Opp-SGO-DE has been used for parameter tuning of prediction models where fitness function is evaluated using the accuracy returned by the prediction models.

Improving Phase

In the improving phase, crossover technique of DE algorithm has been used. Crossover is performed between every individual candidate solution and *'G_{best}'*. Crossover rate *'Cr'* has been used to control the crossover. In every run, a random number is generated and compared to *'C_r'*, if greater, then value from *'G_{best}'* will be copied, else same value will remain. Generally greater value of *'C_r'* is preferred so that only minimum traits of a person or candidate solution are changed. For better understanding, Algorithm-1 can be referred. Flowchart of the same has been depicted in Fig. 1.

Algorithm-1

1. Initialize the population set *'A'* randomly.
2. Evaluate the fitness value *'F_i'* for every A_i, *where* $i= 1,2\ldots n$.
3. Select the *'G_{best}'*, *'C_r' and 'M_r'*.
4. Repeat the following steps till stopping criteria is met.

Improving Phase

5. Select a random number *'r'* which lies between (0,1).
6. For every $A_{(ij)}$, find the *'$A_{imp(ij)}$'* using the following condition-

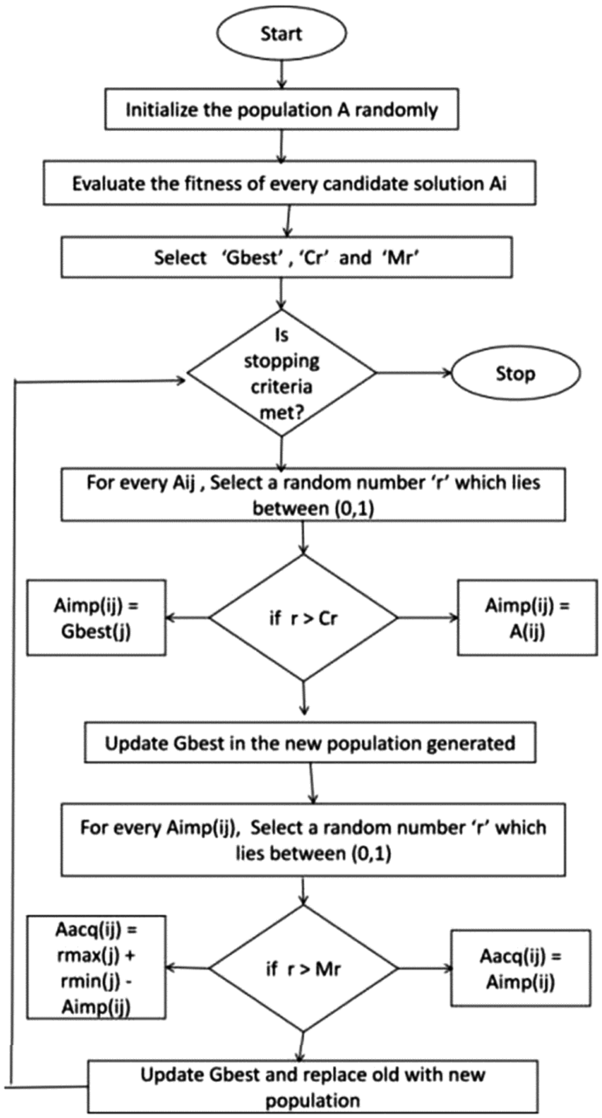

Fig. 1 — Flowchart of Opp-SGO-DE

if $r > C_r$, *then* $A_{imp(ij)} = G_{best(j)}$
else $A_{imp(ij)} = A_{(ij)}$

Acquiring Phase

7. Select a random number *'r'* which lies between (0,1).
8. For every $A_{imp(ij)}$, find the *'*$A_{acq(ij)}$*'* using the following condition-

if $r > M_r$, *then find* $A_{opp(ij)} = r_{max(j)} + r_{min(j)} - A_{imp(ij)}$ *and* $A_{acq(ij)} = A_{opp(ij)}$
else $A_{acq(ij)} = A_{imp(ij)}$

9. Find the fitness of $A_{acq(i\ )}$. If it is better than $A_{(i\ )}$, then replace else discard.
10. Update G_{best} and propagate the updated G_{best} and updated population to the next iteration.

Acquiring Phase

In the acquiring phase, mutation technique of DE algorithm has been used. Mutation is performed for every individual candidate solution and replaced with its *'Opposite Number'*. Mutation rate 'M_r' has been used to control the mutation process. In every run, a random number is generated and compared to *'*M_r*'*, if greater, then its value will be replaced with its opposite number, else same value will remain. Generally greater value of *'*M_r*'* is preferred so that only minimum traits of a person or candidate solution are changed. For better understanding, Algorithm-1 can be referred.

Applying Opp-SGO-DE to SVM and RF for Parameter Tuning

Support Vector Machine (SVM) divides the classes using a hyperplane, used for classification and regression problems. In this paper, a RBF kernel has been used and *'C'*, which deals with mis-classification of training samples and *'γ'*, which deals with the effect of a single training sample, are the two parameters considered for tuning. Random Forest is an ensemble learning technique which learns from multiple trees of random sizes. *'K*' which denotes number of trees and *'M'* which specifies number of variables used to split the nodes, are the two parameters considered for tuning. Accuracy returned by these prediction models is considered as the fitness function when Opp-SGO-DE applied for parameter tuning.

Methodology

All the experiment codes are implemented in MATLAB 2016a. The experiments are conducted on a Intel Core i5, 8 GB memory laptop in Windows 10 environment. In all the simulations, population size was set to 30 and number of iterations was set to 100 and were repeated for 20 times and their average value was considered as the final result.

Dataset

First dataset taken is the 'Indian Trauma Population from Road Traffic Accidents' dataset which was combinedly collected from four public medical colleges across India.[22] Number of samples in the dataset is 15,865 with 119 features. For the classification, only 12 features i.e. ICD-10 codes from emergency department have been considered because in this paper, the main objective is to predict the mortality of a patient from the on-spot RTA features, so that more serious patients could be given emergency care. That's why the feature which shows whether a person will die within 24 hrs or not is taken as the label. Then all the rows having null and NaN values were removed for better accuracy. For better analysis of proposed methodology, it has been simulated with another dataset which details about the 'Total Number of Persons Killed (NPK) due to RTA in Kerala during 2005–2018 recorded at consecutive eight time zones' (https://old.keralapolice.gov.in/public-information/crime-statistics/road-accident/accessed on 20-09-2021). It is a time series data, so to make it suitable for simulating with prediction models, sliding window method has been used. The data consist of 112 observations and were recorded in the consecutive time intervals each of 3h duration such as 00:00 – 03:00, 03:00 – 06:00, 06:00 – 09:00, 09:00 – 12:00, 12:00 – 15.00, 15:00–18:00, 18:00 – 21:00 and 21:00 – 24:00.

Results and Discussion

Opp-SGO-DE was applied to SVM with RBF kernel and RF for parameter tuning. The range of values for 'C' and 'γ' in SVM is [0,1] and range of 'K' in RF is taken as 64-128and 'M' in RF is taken as 1-100. To find the values of 'C_r' and 'M_r' in which SVM and RF yielded best results, their values were set in the range of [0.1,0.9] and simulation was carried out on both the datasets whose results could be observed in Figs 2(a), 2(b), 3(a) and 3(b). While testing for 'C_r', 'M_r' was constantly set to 0.5 and vice versa. Values obtained for 'C_r' and 'M_r' from both the datasets were averaged and considered for simulation. From the observations of Fig. 2 and 3, 'C_r' value is set to 0.75 and 'M_r' value is set to 0.75 in all the simulations of RF and 'C_r' value is set to 0.7 and 'M_r' value is set to 0.6 in all the simulations of SVM. All the hyper-parameter values taken for simulations are summarized in Table 1. Parameter tuning using Opp-SGO-DE helped in increasing the value of Accuracy and F1-Score of RF and SVM, which can be observed in Figs 4 and 5. The convergence graph of RF has been shown in Fig. 4(a),

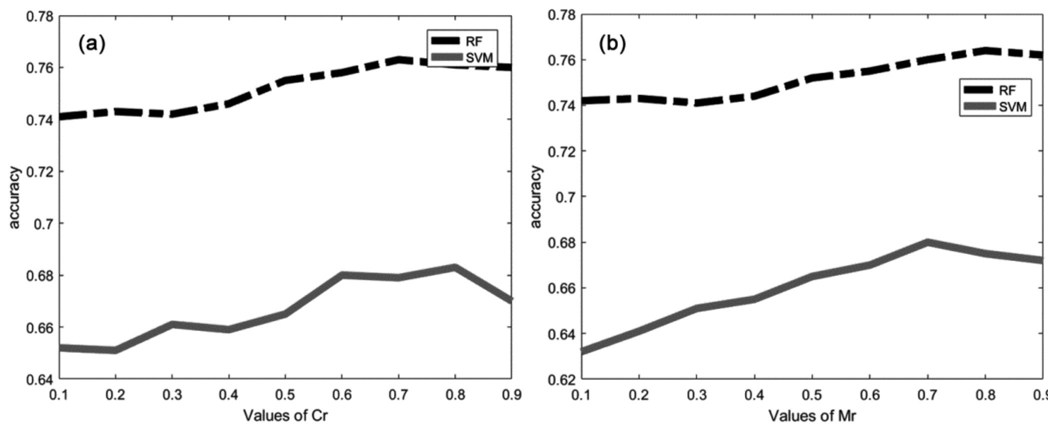

Fig. 2 — (a) Accuracy of RF and SVM plotted against different values of 'C_r' for dataset-1; (b) Accuracy of RF and SVM plotted against different values of 'M_r' for dataset-1

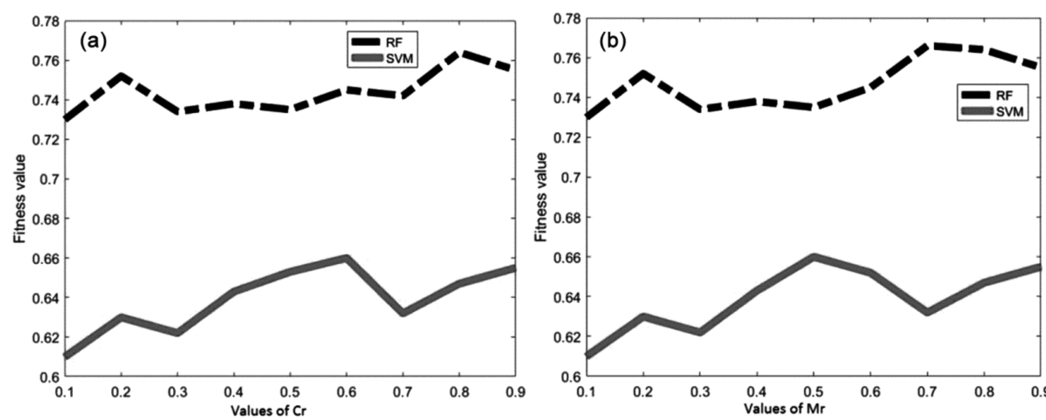

Fig. 3 — (a) Accuracy of RF and SVM plotted against different values of 'C_r' for dataset-2 (b) Accuracy of RF and SVM plotted against different values of 'M_r' for dataset-2

where the accuracy was increased from 0.75 to 0.82 and F1-Score was increased from 0.66 to 0.77 for dataset-1 and Fig. 4(b) shows the convergence graph of SVM where the accuracy was increased from 0.63 to 0.74 and F1-Score was increased from 0.58 to 0.67 for dataset-1. Similarly, Fig. 5(a) shows the convergence graph of RF where the accuracy was increased from 0.66 to 0.75 and F1-Score was increased from 0.62 to 0.65 for dataset-2 and Fig. 4(b) shows the convergence graph of SVM where the accuracy was increased from 0.56 to 0.62 and F1-Score was increased from 0.54 to 0.575 for dataset-2.

Thus it can be clearly deduced that the proposed methodology yields promising results with both RF and SVM as shown in Figs 6 and 7. In RF, the best values for 'K' and 'M' was found to be 100 and 75 respectively for dataset-1. In SVM, the best values for 'C' and 'γ' was found to be 0.156 and 0.967 respectively for dataset-1. In RF, the best values for 'K' and 'M' was found to be 96 and 67 respectively for dataset-2. In SVM, the best values for 'C' and 'γ' was found to be 0.144 and 0.931 respectively for dataset-2. In the literature no work on application of RF and SVM could be found for the used datasets, hence comparative analysis with other such works could not be performed but comparative performance

Table 1 — Values of Hyper-parameters used for simulations

Hyper-parameters	Values
No. Of Population	30
No. Of Iteration	100
'C_r' for RF	0.75
'C_r' for SVM	0.7
'M_r' for RF	0.75
'M_r' for SVM	0.6
Range of 'K' in RF	64-128
Range of 'M' in RF	1-100
Range of 'C' & 'γ' in SVM	[0,1]

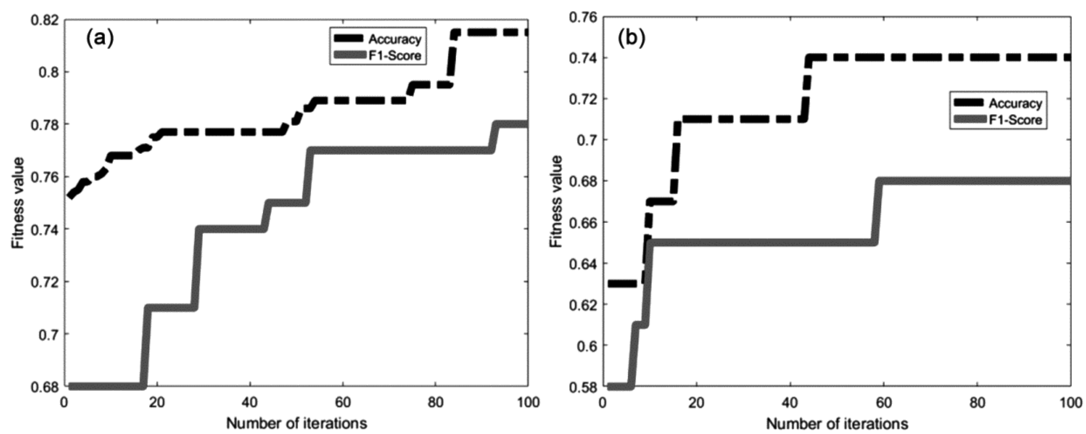

Fig. 4 — (a) Convergence graph of Accuracy and F1-Score for RF for dataset-1; (b) Convergence graph of Accuracy and F1-Score for SVM for dataset-1

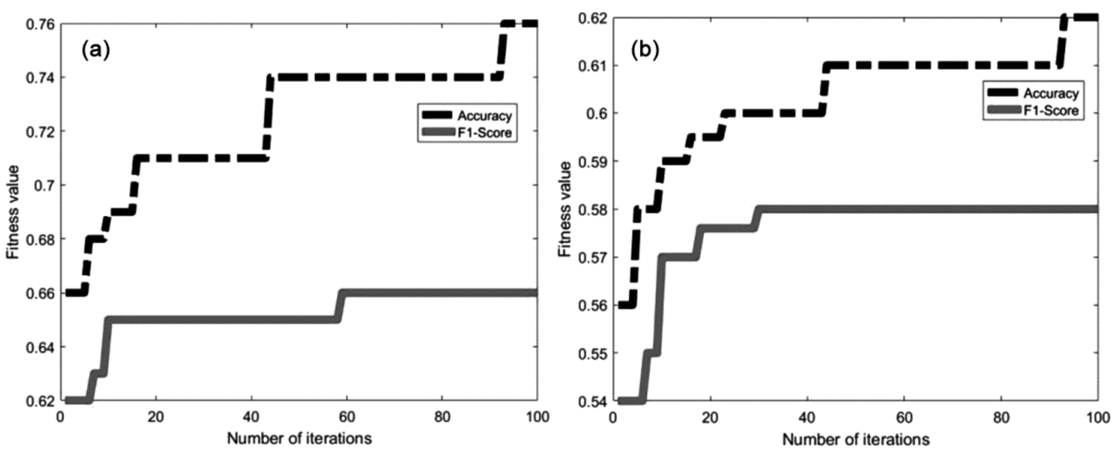

Fig. 5 — (a) Convergence graph of Accuracy and F1-Score for RF for dataset-2; (b) Convergence graph of Accuracy and F1-Score for SVM for dataset-2

Table 2 — Performance analysis of Opp-SGO-DE based RF and SVM

Dataset	Random Forest				Support Vector Machine			
	Accuracy		F1-Score		Accuracy		F1-Score	
	RF	RF-Opp-SGO-DE	RF	RF-Opp-SGO-DE	SVM	SVM-Opp-SGO-DE	SVM	SVM-Opp-SGO-DE
Dataset-1	0.75	0.82	0.66	0.77	0.63	0.74	0.58	0.67
Dataset-2	0.66	0.75	0.62	0.65	0.56	0.62	0.54	0.575

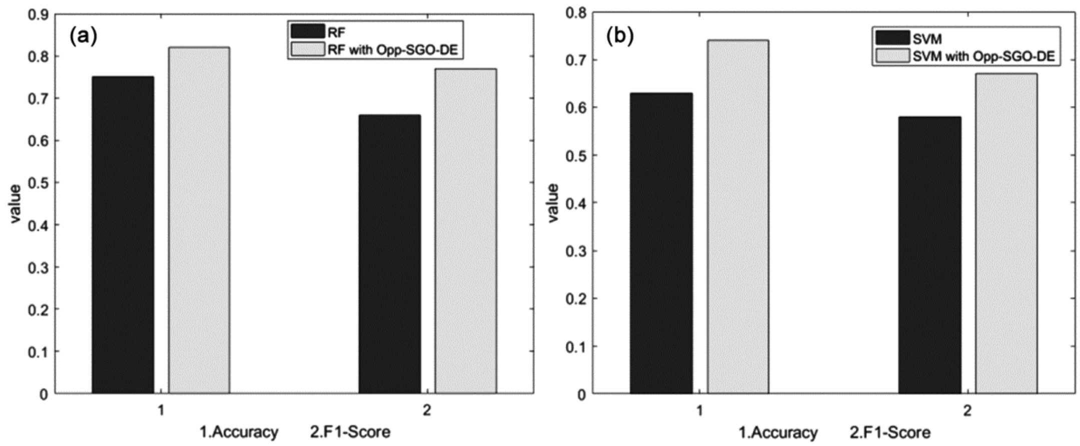

Fig. 6 — (a) Comparison between Accuracy and F1-Score values of RF and RF with Opp-SGO-DE for dataset-1; (b) Comparison between Accuracy and F1-Score values of SVM and SVM with Opp-SGO-DE for dataset-1

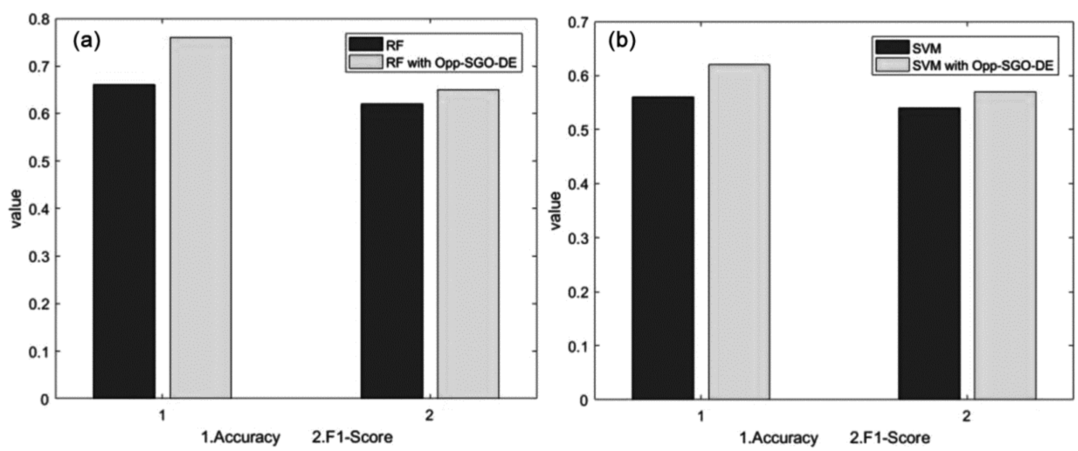

Fig. 7 — (a) Comparison between Accuracy and F1-Score values of RF and RF with Opp-SGO-DE for dataset-2; (b) Comparison between Accuracy and F1-Score values of SVM and SVM with Opp-SGO-DE for dataset-2

analysis of RF and SVM on both the datasets, which is given in Table 2, clearly shows that the proposed Opp-SGO-DE yielded better results and would perform similarly if used with other algorithms too.

Conclusions

The proposed Opp-SGO-DE algorithm provided promising results when applied for hyper-parameter tuning of prediction algorithms such as RF and SVM. A significant increase in Accuracy and F1-score value could clearly be observed from the experiments performed. Number of fitness evaluations have been reduced to half in the proposed algorithm than the original SGO for which significant decrease in time complexity was observed. One limitation of the proposed methodology is increase in algorithmic parameter. In original SGO, only one algorithmic parameter i.e. 'c' or self-instrospection parameter was there, but in proposed methodology two algorithmic parameters i.e. 'C_r' or Crossover rate and 'M_r' or Mutation rate, were introduced but that does not add much to the algorithm's complexity and hence is acceptable. In the literature no work on mortality prediction or analysis using SVM or RF could be found on the used datasets, thus more such works could be carried out and mortality prediction analysis could be performed. This is an emerging domain and could help in development of automated systems for emergency medicines. These type of works, if implemented practically could significantly help developing countries like India where death rate due to RTA is so high.

References

1 World Health Organization, Global status report on road safety 2018, summary, *World Health Organization*, 2018.

2 Boo Y & Choi Y, Comparison of prediction models for mortality related to injuries from road traffic accidents after correcting for undersampling, *Int J Environ Res Public Health*, **18(11)** (2021) 5604.

3 Alharbi R J, Lewis V & Miller C, A state-of-the-art review of factors that predict mortality among traumatic injury patients following a road traffic crash, *Australas Emerg Care*, 2021, Feb 19.

4 Samad A, Khan I A, Afazal A, Shehbaz L, Nasir S & Asim S J, Predictive value of injury severity score in relation to morbidity and mortality following road traffic accident, *Pak J Surg*, **37(2)** (2021) 65–69.

5 Kenneth G E, Statistical application of regression techniques in modeling road accidents in Edo State, Nigeria, *Sch J Phys Math Stat*, **8(1)** (2021) 14–18.

6 Vipin N & Rahul T, Road traffic accident mortality analysis based on time of occurrence: Evidence from Kerala, India, *Clin Epidemiol Glob Health*, **11** (2021) 100745.

7 Satapathy S & Naik A, Social group optimization (SGO): a new population evolutionary optimization technique, *Complex intell*, **2(3)** (2016) 173–203.

8 Naik A, Satapathy S C, Ashour A S & Dey N, Social group optimization for global optimization of multimodal functions and data clustering problems, *Neural Comput Appl* **30(1)** (2018) 271–287.

9 Praveen S P, Rao K T & Janakiramaiah B, Effective allocation of resources and task scheduling in cloud environment using social group optimization, *Arab J Sci Eng*, **43(8)** (2018) 4265–4272.

10 Dey N, Rajinikanth V, Ashour A S & Tavares J M, Social group optimization supported segmentation and evaluation of skin melanoma images, *Symmetry*, **10(2)** (2018) 51.

11 Rajinikanth V & Satapathy S C, Segmentation of ischemic stroke lesion in brain MRI based on social group optimization and fuzzy-Tsallis entropy, *Arab J Sci Eng* (Springer Science & Business Media BV), **43(8)** (2018) 4365–4378.

12 Naik A, Satapathy SC & Abraham A, Modified Social Group Optimization—A meta-heuristic algorithm to solve short-term hydrothermal scheduling, *Appl Soft Comput*, **95** (2020) 106524.

13 Dey N, Rajinikanth V, Shi F, Tavares J M, Moraru L, Karthik K A, Lin H Kamalanand K & Emmanuel C, Social-Group-Optimization based tumor evaluation tool for clinical brain MRI of Flair/diffusion-weighted modality, *Biocybern Biomed Eng*, **39(3)** (2019) 843–856.

14 Dey N, Rajinikanth V, Fong S J, Kaiser M S & Mahmud M, Social group optimization–assisted Kapur’s entropy and morphological segmentation for automated detection of COVID-19 infection from computed tomography images, *Cognit Comput*, **12(5)** (2020) 1011–1023.

15 Chakravarthy V S, Chowdary P S, Satpathy S C, Terlapu S K & Anguera J, Antenna array synthesis using social group optimization, in *Microelectronics, Electromagnetics and Telecommunications* 2018 (pp. 895–905), Springer, Singapore.

16 Fang J, Zheng H, Liu J, Zhao J, Zhang Y & Wang K, A transformer fault diagnosis model using an optimal hybrid dissolved gas analysis features subset with improved social group optimization-support vector machine classifier, *Energies*, **11(8)** (2018) 1922.

17 Jena J J & Satapathy S C, A new adaptive tuned Social Group Optimization (SGO) algorithm with sigmoid-adaptive inertia weight for solving engineering design problems, *Multimed Tools Appl*, **24** (2021) 1–35.

18 Verma S, Jena J J, Satapathy S C & Rout M, Solving travelling salesman problem using discreet social group optimization, *J Sci Ind Res*, **79(10)** (2020) 928–930.

19 Das S, Saha P, Satapathy S C, Jena J J, Social group optimization algorithm for civil engineering structural health monitoring, *Eng Optim*, **53(10)** (2021) 1651–1670, DOI: 10.1080/0305215X.2020.1808974

20 Karaboğa D & Ökdem S, A simple and global optimization algorithm for engineering problems: differential evolution algorithm, *Turk J Elec Eng & Comp Sci*, **12(1)** (2004) 53–60.

21 Rahnamayan S, Tizhoosh H R & Salama M M, Opposition versus randomness in soft computing techniques, *Appl Soft Comput*, **8(2)** (2008) 906–918.

22 Attergrim J, Sterner M, Claeson A, Dharap S, Gupta A, Khajanchi M, Kumar V & Gerdin Wärnberg M, Predicting mortality with the international classification of disease injury severity score using survival risk ratios derived from an Indian trauma population: A cohort study, *PloS one*, **13(6)** (2018) e0199754.