International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

Optimizing Road Transportation Big Data: A Novel Approach for Feature Selection through Optimization Techniques

B. Sangeetha[1], Dr. M. Prabakaran[2]
[1]Research Scholar, PG and Research Department of Computer Science, Government Arts College (Autonomous) (Affiliated to Bharathidasan University, Tiruchirappalli), Karur -5, Tamilnadu, India.
[2]Research Advisor& Associate Professor, PG and Research Department of Computer Science, Government Arts College (Autonomous) (Affiliated to Bharathidasan University, Tiruchirappalli), Karur – 5, Tamilnadu, India.

Abstract

Road traffic accidents are very essential for common people, consequential an estimated 1.2 million deaths and 50 million injuries all over the world every year. In this emerging world, the road accidents are among the principal reason of fatality and injury. The concern of traffic safety has heaved immense alarms across the manageable enhancement of contemporary traffic and transportation. The analysis on road traffic accident grounds can detect the major aspects quickly, professionally and afford instructional techniques to the prevention of traffic accidents and reduction of road traffic accident, which might significantly decrease personal victim by means of road traffic accidents. Data Mining techniques are used in the process of knowledge discovery for many domains' problems. Feature Selection plays a vital role for a large number of datasets. In this research work, the model involves two main phases (i) Feature Selection and (ii) Classification. Since the length of feature vector tends to high, optimal feature selection technology is included, from which the most relevant features are selected by the Lion-based Firefly Algorithm which is referred as Optimization based Feature Selection Method (OFSM). The main objective of this paper is projected on minimizing the correlation between the selected features, which results in providing diverse information regarding the different classes of data. Once, the optimal features are selected, the classification algorithm called Neural Network (NN) is adopted, which can classify the data in an effective manner with the selected features.

Keywords: Feature Selection, Classification, Optimization algorithms, Artificial Neural Network, Transportation, Fire Fly Optimization, Animal Migration Optimization.

1. Introduction

India has the second biggest road network in the world. Therefore, road accidents occur rather recurrently and moreover, they assert several lives every year. It is essential to detect the origin for road accidents with the aim of evade them. In the recent years the compilation of data on traffic amount has turned out to be a huge part of the work of road forecasting programs by both value and workforce. The Traffic data is apportioned into various rules by categorizing the breakpoints for traffic factors in the data. In two-regime traffic methods, decisive possession is employed to divide free flow and crammed flow conditions. Traffic clogging roots incredible failure in terms of both time and power worn out. The forecast of Traffic flow is an essential detective subject in an Intelligent Transportation System (ITS) [1], and it can be utilized as an vital measurement to sort the predicament of traffic congestion. Traffic clogging is rooted while the traffic stipulates methods or surpasses the obtainable facility of the traffic network. Fortunately, as a result of systematic sensor instrumentations of road networks in main metropolis in addition to the vast accessibility of supplementary commodity sensors from which traffic data can be copied (e.g., CCTV cameras, GPS devices), a huge amount of instantaneous and chronological traffic data at very high spatial and sequential decisions has grown to be obtainable [2].

The concern of affording the travel safety on the road network in the urban and suburban is one of the primary ethics which leads the engineering, traffic and transportation development. Approximately 3,500 people die by the road accidents all over the world every day. Tens of millions of people are wounded or disabled each year. Pedestrians, children, cyclists, bike, car users and the aged are amongst the most susceptible of road users. WHO functions with partners - governmental and nongovernmental - around the globe to elevate the report of the inevitability of road traffic injuries and support high-quality trainings associated to helmet and seat-belt wearing, drunk and drive, rash riding and being noticeable in traffic [2]. On the other hand, the accident is predictable, specified the definition, "an incident is distinct as the rapid accidental release of or the spotlight to a dangerous substance that marks in or might logically have brought about the injuries, deaths, considerable possessions or ecological harm, evacuation or sheltering in place [3]". The expense of these accidents can be serious trouble to the government. Therefore, the road accidents are a severe threat to public transportation. In particular, the colossal economic toll of road accidents on human societies inflict, recuperating road safety needs awareness to three effective features: human, the road and the vehicle [4] [5]. The bureaucrats of the transport system also require stabling the road safety wants through inadequate resources to lessen the accidents and augment the road conditions. In fact, the p is as primary objective is as conceivable in reducing accidents that can be loomed to this target by managing the traffic engineering, driver training and implementation [6].

2. Related Works

Kong, Xiangjie, et al [7] proposed LoTAD to discover abnormal areas with long-standing poor traffic conditions. In particular, the authors methods pack sourced bus data into TS-segments (Temporal and Spatial segments) to form the traffic condition. In a while, the authors investigate anomalous TS-segments in each bus line by scheming their AI (Anomaly Index). Subsequently, the authors merge anomalous TS-segments identified in various lines to mine anomalous regions. The data of anomalous areas affords recommendations for prospect traffic planning.

Kumar, PriyanMalarvizhi, et al [8] suggested an valuable traffic control method with the support of IoV technology. The proposed technique is established in the study area of Vellore district, Tamil Nadu, India. The street maps are segmented into small number of different maps. Ant colony algorithm is practiced to each map with the intention to discover the optimal route. In addition, Fuzzy logic-based traffic intensity calculation function is proposed in this paper to form the heavy traffic. The planned IoV based route selection technique is contrasted with the active shortest path selection algorithms like Prim's algorithm, Kruskal's algorithm, and Dijikstra algorithm.

Lin, Lu, et al [9] intended to advance the road traffic speed prediction by combining the customary speed sensing data with new-type "sensing" data from cross domain sources, such as tweet sensors from social media and route sensors from map and traffic service platforms. Jointly modelling data from numerous datasets fetches several disputes, including location uncertainty of low-resolution data, language vagueness of traffic depiction in texts, and heterogeneity of cross-domain data. Due to these difficulties, the authors exhibit a unified probabilistic framework, called Topic-Enhanced Gaussian Process Aggregation Model (TEGPAM).

Moriya, Koichi, Shin Matsushima, and Kenji Yamanishi [10] proposed a new method for mining traffic menace from such assorted data. Traffic threat refers to the prospect of event of traffic accidents.

Zhang, Zhenhua, et al [11] occupied deep learning in identifying the traffic accident from social media data. Initially, the authors methodically explore the 1-year over 3 million tweet contents in two metropolitan regions: Northern Virginia and New York City. The paired tokens can incarcerate the association rules inbuilt in the accident-related tweets and enhance the precision of the traffic accident detection. Second, two deep learning methods: Long Short-Term Memory (LSTM) and Deep Belief Network (DBN) are examined and employed on the extorted token.

3. Proposed Optimization based Feature Selection Method for Classification

The selection of an optimal feature subset from the high-dimensional feature set is a critical task in the big data mining as it involves the collection and processing of vast amounts of data [12][13]. The classical and advanced data mining and machine learning tools that are available in the present trend are not sufficient to extract the features in an optimal way. Since, the length of the features is found to be larger using there is a necessity for an optimal feature selection technology [14][15], which is accomplished by the proposed hybrid model with the combination of FF and AMO called AMO+FF that results in selecting the most relevant features. Moreover, feature classification is a decisional approach in which the features of the data are grouped under a certain class that is specified in priori. Once, the optimal feature is selected, the classification algorithm called NN is adopted, which can classify the data in an effective manner with the selected features.

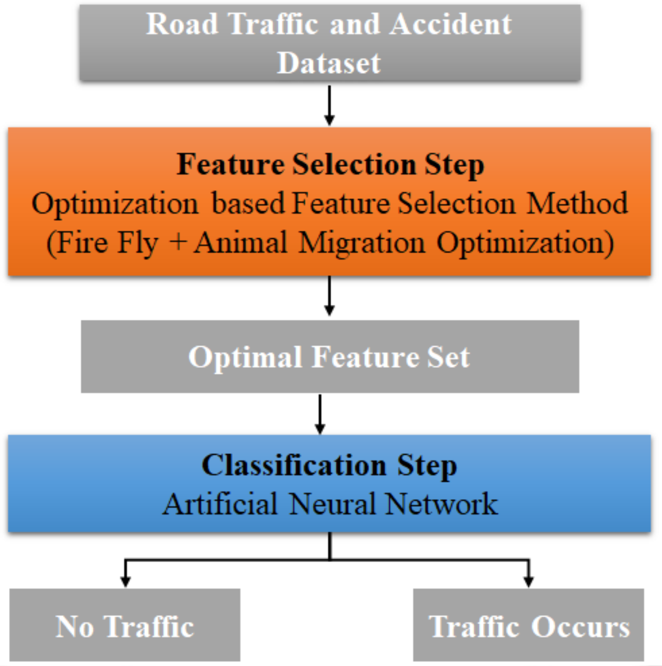

Figure 1: Proposed Optimization based Feature Selection Method for Classification

3.1 Firefly Algorithm

FF algorithm [16] was developed by Xin-She Yang in the year 2008 by the inspiration got from the fireflies. Three main assumptions were made here; they are (a) all FF are unisex (b) Attractiveness is directly proportional to brightness and attractiveness is inversely proportionally to distance. (c) The objective function defines the brightness of FF. Each FF has its attractiveness which is represented as ρ and it decreased with distance x. Equation (2) represents the attractiveness between two FF in which ρ_0 denotes the maximum attractiveness and it is referred as the light absorption coefficient. Further, g and h are the two FF at position K_g and K_h, their distance is evaluated using the mathematical equation (3) in which b represents the count of dimensions. The movement of FF is represented in Eq. (4). The light intensity M_h of FF is evaluated on the basis of the distance between the fireflies. The mathematical equation of FF is shown in Eq. (1) in which M_0 represents the original light intensity [17].

$$M = M_0 e^{-x} \qquad (1)$$

$$\rho(x) = \rho_0 e^{-x}, v \geq 1 \qquad (2)$$

$$x_{gh} = \|K_g - K_h\| = \sqrt{\sum_{w=1}^{b}(K_{g,w} - K_{h,w})^2} \qquad (3)$$

$$K_{best} = K_g + \rho_0^{-\gamma x_{gh}^2}(K_h - K_g) + \omega\left(rand - \frac{1}{2}\right) \qquad (4)$$

The current position of FF is denoted by the first term and the attractiveness of FF is denoted by the second term. The random movement of FF is described by the last term. The initial position of FF is denoted as per Eq. (4). The pseudo code for conventional FF is shown in Algorithm 1.

Algorithm 1: Firefly Algorithm

Step 1: Initialize Maximum generation Max_g and intensity of light M_g Light Absorption co-efficient is defined.

Step 2: While (t<Max_g)

Step 3: For g=1: n_1 for all FF.

Step 4: For h=1: n_2 for all FF

Step 5: IF ($M_h > M_g$)

Step 6: FF g is moved towards h

Step 7: End if

Step 8: Attractiveness varies with distance x.

Step 9: New solutions are evaluated and light intensity is updated

Step 10: End for h

Step 11: End for g

Step 12: FF are ranked and the best FF is predicted

Step 13: End while

3.2 Animal Migration Optimization (AMO) Algorithm

The AMO calculation is a swarm-based calculation roused by the migration wonder of animals [18][19]. In the calculation, people are viewed as places of animals, and positions can be moved by for the most part two activities: creature migration and populace refreshing. The activity of creature migration reproduces practices of creature bunches moving from the present territory to another region. New places of people will be delivered by the bearing of creature migration, where three migration rules are considered: Individuals move towards a similar course as their neighboring people; people stay close to their neighboring people; and people keep away from crashes with their neighboring people. Utilizing the three migration runs, a probability approach is acquainted with yield new places of people. The calculation starts with a haphazardly instatement populace, which is contained NP include vectors with D_x Dimensions, which can be expressed as pursues:

$$y_{j,k,0} = y_{k,min} + rand_{j,k}.(y_{k,max} - y_{k,min}) \quad (5)$$

Where $y_{k,max}$ and $y_{k,min}$ are the maximum value and minimum value of the k-ith dimension. $y_{j,k,0}$ is the k-th dimension value of the j-th individual in the initialization population, and $rand_{j,k}$ is a uniformly random number between 0 and 1, j=1,…,NP and k=1,…, D_x.

In the wake of delivering the introduction populace, animal migration and populace refreshing tasks are performed iteratively. During the animal migration, the people should move another situation as indicated by the places of their neighboring people, which can be depicted as pursues:

$$y_{j,k,G+1} = y_{j,k.G} + \delta.(y_{neighborhood,k,G} - y_{j,k,G}) \quad (6)$$

Where $y_{j,k,G}$ is the k-th dimension value of the j-th individual in the current population G, and $y_{j,k,G+1}$ is the k-th dimension value of the j-th individual of the new population G+1; $y_{neighborhood,k,G}$ is the k-th dimension value of the neighbouring individual of $y_{j,k,G}$ which is defined using a ring topology scheme. For the k-th dimension, $y_{neighborhood,k,G}$ is selected from the (j-2)-th individual, the (j-1)-th individual, the j-th individual, the (j+1)-th individual, and the (j+2)-th individual, δ is a random number produced by a Gaussian Distribution with N (0,1).

The population updating simulates how animals leave the group and new individuals join in the new population, as the equation describes:

$$y_{j,k,G+1} = y_{r1,k.G} + rand_1.\left(y_{best,k,G} - y_{j,k,G}\right) + rand_2.\left(y_{r2,k,G} - y_{j,k,G}\right) \quad (7)$$

Where $y_{r1,k.G}$ is the k-ith dimension value of the individual to be updated, which is chosen randomly in the current population; moreover, different from $y_{j,k,G}, y_{r2,k,G}$ is the k-ith dimension value of another random individual, $y_{best,k,G}$ is the k-ith dimension value of the best individual that has been found. $rand_1$ and $rand_2$ are two uniformly random numbers between 0 and 1. The algorithm makes the assumption that the number of animals in the population remains unchanged. Therefore, in the updating, it replaces some of the animals with new individual according to a probability p_{a_j}, which is related to the fitness of individuals and can be calculated as follows:

$$p_{a_j} = \frac{sn_j}{NP} \quad (8)$$

Where p_{a_j} is the probability value of the j-th individual, NP is the number of the individuals in the population, and sn_j is the sequence number of the fitness of j-th individual after being sorted by their fitness in descending order, where i = 1, 2, . . . , NP.

For every person and each measurement, a consistently arbitrary number, meant by rand, somewhere in the range of 0 and 1 will be created as the probability to decide if the individual is held or is supplanted by another person. In this way, people with better wellness will be saved with higher probability in the people to come, while those with more regrettable wellness will likely be supplanted by new people. Besides, the animal with best position will be held in the people to come.

5.5 Proposed Optimization based Feature Selection Algorithm

In general, the FF has the capacity of dealing complex non-linear, multi-modal optimization problem in an efficient way. Moreover, it does not require a good solution in order to begin the iteration process. Beyond this, the FF suffers from the drawbacks like, the parameters of FF algorithm are set as fixed and they do not get altered with respect to time. FF do not have the capacity of remembering the history of better situation of every other FF, hence there is a chance for missing of the best solution. The firefly has no capacity of migrating in a random behavior. The FF could move to a certain direction alone where the brightness has been enhanced. So, due to this complex nature, FF has been linked with AMO algorithm. The interlinked algorithm improved the quality of the optimal solution. Initially, in the proposed O-FS algorithm, a random number is selected. If the random value is found to be greater than 0.5, the solution is updated with the FF. This renewal of the solution is accomplished with respect to Eq. (4). In case, if the random value is found be higher than 0.5, then the solution is updated by the AMO and the renewal of the best solution is accomplished with respect to Eq. (7). The pseudo code for O-FS algorithm is represented in Algorithm 2.

Algorithm 2: Proposed Optimization based Feature Selection

Step 1: Initialize Maximum generation Max_e and intensity of light M_g Light absorption co-efficient is defined.

Step 2: while (t<Max_g)

Step 3: For g=1:n for all FF

Step 3.1: For h=1:g for all FF

Step 3.2: IF (rand <0.5)

Update the solution using Eq. (4) by FF algorithm

Else

Update the solution using Eq. (7) by AMO

Step 3.3: If ($M_h > M_g$)

Step 3.4: Firefly g is moved towards h

Step 3.5: else

Step 3.6: Attractiveness varies with distance x

Step 3.7: New solutions are evaluated and light intensity is updated

Step 3.8: end

Step 3.9: End if

Step 4: end

Step 5: end

Hence, the optimally generated features are represented as $F_i^* = \{F_1^*, F_2^*, .., F_n^*\}$. The flowchart of the proposed Optimization based Feature Selection model is shown in the figure 3.

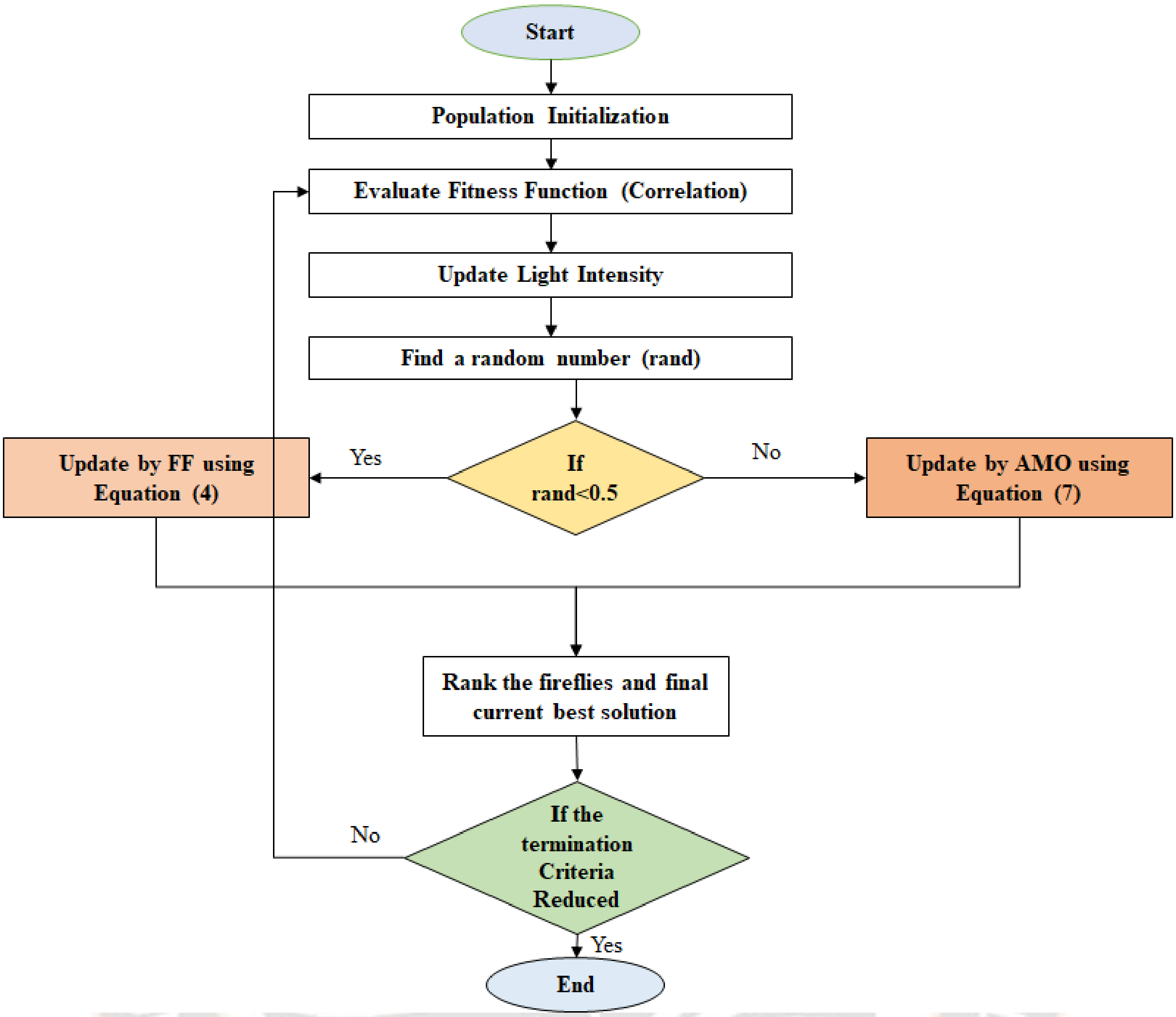

Figure 3: Flowchart of the Proposed Optimization based Feature Selection

4. Result and Discussion

4.1 Dataset Description

The road accident dataset is taken from the Kaggle Repository. Table 1 depicts the features involved in the road accident dataset [R].

Table 1: Description of the Dataset

Feature Number	Feature Name
1	accident_index
2	vehicle_reference
3	vehicle_type
4	towing_and_articulation
5	vehicle_manoeuvre
6	vehicle_location-restricted_lane
7	junction_location
8	skidding_and_overturning
9	hit_object_in_carriageway
10	vehicle_leaving_carriageway
11	hit_object_off_carriageway

12	1st_point_of_impact
13	was_vehicle_left_hand_drive?
14	journey_purpose_of_driver
15	sex_of_driver
16	age_of_driver
17	age_band_of_driver
18	engine_capacity_(cc)
19	propulsion_code
20	age_of_vehicle
21	driver_imd_decile
22	driver_home_area_type
23	vehicle_imd_decile
24	NUmber_of_Casualities_unique_to_accident_index
25	No_of_Vehicles_involved_unique_to_accident_index
26	location_easting_osgr
27	location_northing_osgr
28	longitude
29	latitude
30	police_force
31	accident_severity
32	number_of_vehicles
33	number_of_casualties
34	date
35	day_of_week
36	time
37	local_authority_(district)
38	local_authority_(highway)
39	1st_road_class
40	1st_road_number
41	road_type
42	speed_limit
43	junction_detail
44	junction_control
45	2nd_road_class
46	2nd_road_number
47	pedestrian_crossing-human_control
48	pedestrian_crossing-physical_facilities
49	light_conditions
50	weather_conditions
51	road_surface_conditions
52	special_conditions_at_site
53	carriageway_hazards
54	urban_or_rural_area
55	did_police_officer_attend_scene_of_accident
56	lsoa_of_accident_location
57	casualty_reference
58	casualty_class
59	sex_of_casualty
60	age_of_casualty
61	age_band_of_casualty

62	casualty_severity
63	pedestrian_location
64	pedestrian_movement
65	car_passenger
66	bus_or_coach_passenger
67	pedestrian_road_maintenance_worker
68	casualty_type
69	casualty_home_area_type
70	casualty_imd_decile

4.2 Number of Features obtained

Table 2 depicts the number of features obtained by the Proposed Optimization based Feature Selection Algorithm, FireFly algorithm, Animal Migration Optimization algorithm. From the table 2, it is clear that the proposed optimization-based feature selection method gives less number of features than the existing optimization algorithm like Firefly algorithm and Animal Migration Optimization algorithm. In the table 2, Proposed optimization-based feature selection algorithm gives only 30 features, FireFly algorithm gives 34 features whereas Animal Migration Optimization algorithm gives 42 features.

Table 2: Number of Features obtained by the Proposed Optimization based Feature Selection Method, FireFly Algorithm and Animal Migration Optimization algorithm

S.No	**Feature Selection Techniques**		
	Proposed Optimization based Feature Selection	**FireFly Algorithm**	**Animal Migration Optimization algorithm**
1	vehicle_type	casualty_home_area_type	accident_severity
2	towing_and_articulation	casualty_type	number_of_vehicles
3	vehicle_location-restricted_lane	pedestrian_road_maintenance_worker	number_of_casualties
4	junction_location	weather_conditions	date
5	skidding_and_overturning	2nd_road_class	day_of_week
6	hit_object_in_carriageway	was_vehicle_left_hand_drive?	time
7	vehicle_leaving_carriageway	journey_purpose_of_driver	local_authority_(district)
8	hit_object_off_carriageway	junction_location	local_authority_(highway)
9	was_vehicle_left_hand_drive?	vehicle_location-restricted_lane	1^{st}_road_class
10	sex_of_driver	vehicle_manoeuvre	1^{st}_road_number
11	age_of_driver	accident_index	road_type
12	engine_capacity_(cc)	hit_object_off_carriageway	road_type
13	age_of_vehicle	vehicle_leaving_carriageway	speed_limit
14	NUmber_of_Casualities_unique_to_accident_index	hit_object_in_carriageway	junction_detail
15	No_of_Vehicles_involved_unique_to_accident_index	age_of_driver	junction_control
16	longitude	age_band_of_driver	road_surface_conditions
17	latitude	engine_capacity_(cc)	special_conditions_at_site
18	accident_severity	propulsion_code	carriageway_hazards
19	number_of_vehicles	age_of_vehicle	urban_or_rural_area

20	number_of_casualties	driver_imd_decile	did_police_officer_attend_scene_of_accident
21	date	driver_home_area_type	lsoa_of_accident_location
22	day_of_week	vehicle_imd_decile	accident_index
22	time	location_easting_osgr	vehicle_reference
23	local_authority_(district)	location_northing_osgr	vehicle_type
24	local_authority_(highway)	longitude	towing_and_articulation
25	road_type	latitude	vehicle_manoeuvre
26	speed_limit	number_of_vehicles	vehicle_location-restricted_lane
27	weather_conditions	number_of_casualties	junction_location
28	road_surface_conditions	date	was_vehicle_left_hand_drive?
29	casualty_severity	day_of_week	journey_purpose_of_driver
30	car_passenger	time	sex_of_driver
31	bus_or_coach_passenger	local_authority_(district)	age_of_driver
32		local_authority_(highway)	age_band_of_driver
33		road_surface_conditions	engine_capacity_(cc)
34		casualty_severity	propulsion_code
35			age_of_vehicle
36			driver_imd_decile
37			NUmber_of_Casualities_unique_to_accident_index
38			No_of_Vehicles_involved_unique_to_accident_index
39			location_easting_osgr
40			location_northing_osgr
41			longitude
42			latitude

4.3 Performance Analysis of the Feature Selection Methods

The classification techniques like Artificial Neural Network, Support Vector Machine and Naïve Bayes techniques are considered to evaluate the performance analysis of the proposed optimization-based feature selection method, firefly algorithm and Animal Migration Optimization algorithm. Table 3 depicts the performance metrics that are considered in this study for the evaluation.

Table 3: Performance Metrics for the evaluation of the proposed optimization-based feature selection method

Metrics	Equation
Accuracy	$\frac{TP+TN}{TP+FN+TN+FP}$
True Positive Rate (TPR)	$\frac{TP}{TP+FN}$
False Positive Rate (FPR)	$\frac{FP}{FP+TN}$
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$
F-Measure	$2.\frac{Precision.Recall}{Precision+Recall}$

Mean Absolute Error (MAE)	$\frac{1}{N}\sum_{I=1}^{N}\left\lvert\hat{\theta}_i-\theta_i\right\rvert$
Root Mean Squared Error (RMSE)	$\sqrt{\frac{\mathbf{1}}{\boldsymbol{N}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}}\left(\hat{\theta}_i-\theta_i\right)^{\mathbf{2}}}$
Relative Absolute Error (RAE)	$\frac{\sum_{I=1}^{N}\left\lvert\hat{\theta}_i-\theta_i\right\rvert}{\sum_{I=1}^{N}\left\lvert\bar{\theta}_i-\theta_i\right\rvert}$
Root Relative Squared Error (RRSE)	$\sqrt{\frac{\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}}\left(\hat{\theta}_i-\theta_i\right)^{\mathbf{2}}}{\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}}(\bar{\theta}_i-\theta_i)^{\mathbf{2}}}}$

Table 4 depicts the performance analysis of the accuracy of the existing and proposed feature selection method using ANN, NB and SVM classification techniques. From the table 4, it is clear that the SVM gives maximum accuracy for original dataset and AMO, FFA processed dataset and proposed feature selection processed dataset gives maximum accuracy when using Artificial Neural Network as the classifier.

Table 4: Performance analysis on the accuracy of proposed OFSM, FFA and AMO feature selection methods

Feature Selection Methods	**Feature Selection Techniques**		
	ANN	**NB**	**SVM**
Original Dataset	51.667%	49.333%	**73.667%**
FFA processed dataset	**73.667%**	50.333%	69.667%
AMO processed dataset	73.333%	55.333%	**87.333%**
Proposed OFSM processed dataset	**92.3667%**	70.566%	88.233%

Table 5 depicts the performance analysis on Kappa Statistic of the proposed OFSM and existing feature selection methods using ANN, NB and SVM classification methods. From the table 5, it is clear that the ANN gives maximum value of Kappa Statistics for original dataset and proposed OFSM processed dataset. SVM gives the maximum Kappa Statistic value for FFA processed dataset and AMO processed datasets gives maximum Kappa statistic value using NB.

Table 5: Performance analysis on Kappa Statistic of the proposed OFSM and existing feature selection methods using ANN, NB and SVM

Feature Selection Methods	**Feature Selection Techniques**		
	ANN	**NB**	**SVM**
Original Dataset	**0.1465**	0.0397	0
FFA processed dataset	0	0.0452	**0.0471**
AMO processed dataset	0.0066	**0.0339**	0.0321
Proposed OFSM processed dataset	**0.712**	0.624	0.601

Table 6 depicts the performance analysis on True Positive Rate (TPR) of the proposed OFSM method and existing feature selection methods using ANN, NB and SVM classification methods. From the table 6, ANN gives maximum value of True Positive Rate (TPR) for FFA processed dataset and proposed OFSM method processed dataset, Original dataset and AMO processed datasets gives maximum TPR value using SVM.

Table 6: Performance analysis on True Positive Rate (Recall) of the proposed OFSM and existing feature selection methods using ANN, NB and SVM

Feature Selection Methods	**Feature Selection Techniques**		
	ANN	**NB**	**SVM**
Original Dataset	0.517	0.493	**0.737**

FFA processed dataset	**0.737**	0.5	0.697
AMO processed dataset	0.733	0.553	**0.873**
Proposed OFSM processed dataset	**0.892**	0.712	0.881

Table 7 depicts the performance analysis on False Positive Rate (RPR) of the proposed OFSM method and existing feature selection methods using ANN, NB and SVM classification methods. From the table 7, ANN gives minimum value of False Positive Rate (FPR) for FFA, AMO processed datasets and proposed OFSM processed dataset and original dataset gives least FPR using NB.

Table 7: Performance analysis on False Positive Rate (FPR) of the proposed OFSM and existing feature selection methods using ANN, NB and SVM

Feature Selection Methods	Feature Selection Techniques		
	ANN	**NB**	**SVM**
Original Dataset	0.741	**0.696**	0.737
FFA processed dataset	**0.439**	0.539	0.702
AMO processed dataset	**0.534**	0.581	0.854
Proposed OFSM processed dataset	**0.112**	0.421	0.582

Table 8 depicts the performance analysis on Precision of the proposed OFSM method and existing feature selection methods using ANN, NB and SVM classification methods. From the table 8, ANN gives maximum value of precision for original dataset, AMO and Proposed OFSM processed datasets and FFA processed datasets gives least Precision value using NB.

Table 8: Performance analysis on Precision of the proposed OFSM and existing feature selection methods using ANN, NB and SVM

Feature Selection Methods	Feature Selection Techniques		
	ANN	**NB**	**SVM**
Original Dataset	**0.543**	0.493	0.488
FFA processed dataset	0.608	**0.637**	0.543
AMO processed dataset	**0.812**	0.798	0.542
Proposed OFSM processed dataset	**0.896**	0.821	0.833

Table 9 depicts the performance analysis on F-Measure of the proposed OFSM method and existing feature selection methods using ANN, NB and SVM classification methods. From the table 9, ANN gives maximum value of F-Measure only for the proposed OFSM method, whereas other datasets give maximum value when using SVM as the classifier.

Table 9: Performance analysis on F-Measure of the proposed OFSM and existing feature selection methods using ANN, NB and SVM

Feature Selection Methods	Feature Selection Techniques		
	ANN	**NB**	**SVM**
Original Dataset	0.502	0.492	**0.640**
FFA processed dataset	0.640	0.560	**0.649**
AMO processed dataset	0.623	0.653	**0.841**
Proposed OFSM processed dataset	**0.893**	0.762	0.826

Table 10 presents the performance analysis on Mean Absolute Error (MAE) of the proposed OFSM method and existing feature selection methods using ANN, NB and SVM classification methods. From the table 10, ANN gives least MAE value for original dataset and proposed OFSM method processed dataset, whereas other feature selection processed datasets give minimum value when using SVM as the classifier.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

Table 10: Performance analysis on Mean Absolute Error (MAE) of the proposed OFSM and existing feature selection methods using ANN, NB and SVM

Feature Selection Methods	**Feature Selection Techniques**		
	ANN	**NB**	**SVM**
Original Dataset	**0.3358**	0.5201	0.6692
FFA processed dataset	0.2633	0.5044	**0.2033**
AMO processed dataset	0.2667	0.4459	**0.1267**
Proposed OFSM processed dataset	**0.0536**	0.1357	0.0891

Table 11 presents the performance analysis on Root Mean Squared Error (RMSE) of the proposed OFSM method and existing feature selection methods using ANN, NB and SVM classification methods. From the table 11, ANN gives least RMSE value for FFA, and proposed OFSM method processed dataset, original dataset gives least RMSE value when using NB as the classifier. The other processed dataset gives least RMSE value using SVM.

Table 11: Performance analysis on Root Mean Squared Error (RMSE) of the proposed OFSM and existing feature selection methods using ANN, NB and SVM

Feature Selection Methods	**Feature Selection Techniques**		
	ANN	**NB**	**SVM**
Original Dataset	0.6892	**0.6301**	0.6845
FFA processed dataset	**0.5132**	0.6	0.6508
AMO processed dataset	0.5164	0.544	**0.4559**
Proposed OFSM processed dataset	**0.1641**	0.4245	0.3614

Table 12 presents the performance analysis on Relative Absolute Error (RAE) of the proposed OFSM method and existing feature selection methods using ANN, NB and SVM classification methods. From the table 12, ANN gives least RAE value for original, FFA, and proposed OFSM method processed dataset, whereas AMO processed dataset gives least RAE value when using SVM as the classifier.

Table 12: Performance analysis on Relative Absolute Error (RAE) of the proposed OFSM and existing feature selection methods using ANN, NB and SVM

Feature Selection Methods	**Feature Selection Techniques**		
	ANN	**NB**	**SVM**
Original Dataset	**96.619**	238.1586	186.7589
FFA processed dataset	**67.7257**	129.8562	78.0131
AMO processed dataset	68.583	204.1779	**58.0022**
Proposed OFSM processed dataset	**51.811**	102.8178	57.2112

Table 13 presents the performance analysis on Root Relative Squared Error (RRSE) of the proposed OFSM method and existing feature selection methods using ANN, NB and SVM classification methods. From the table 13, ANN gives least RRSE value for FFA, and proposed OFSM method processed dataset, whereas original, AMO processed dataset gives least RAE value when using SVM as the classifier.

Table 13: Performance analysis on Root Relative Squared Error (RRSE) of the proposed OFSM and existing feature selection methods using ANN, NB and SVM

Feature Selection Methods	**Feature Selection Techniques**		
	ANN	**NB**	**SVM**
Original Dataset	145.1707	191.5635	**141.5383**
FFA processed dataset	**116.5024**	136.2211	125.0381
AMO processed dataset	112.2375	165.4063	**87.2055**
Proposed OFSM processed dataset	**67.1468**	80.7182	75.1866

5. Conclusion

In this paper, a new classification model was developed with the assistance got from the intelligent methods. Three main phases were used in this paper; they are (i) Optimal Feature selection and (iii) Classification. The dimensions of the data were high, and hence the selection of the optimal features was a complex task so that the proposed model used the optimal feature selection technology referred as OFSM to select the optimal features. The research was focused on the objective of diminishing the correlation between the selected features, as they were related to the generation of diverse information that was related to different classes of data. Further, the classification of the feature was carried out after the selection of the optimal features. The classification of the selected features was done with NN, SVM and NB which had the ability to classify the data in an effective manner with the selected features. Thus, the entire experimental analysis confirms the effective performance of proposed OFSM method for the road accident dataset classification method.

REFERENCE

[1] Alsolami, Bdoor, Rashid Mehmood, and Aiiad Albeshri. "Hybrid Statistical and Machine Learning Methods for Road Traffic Prediction: A Review and Tutorial." Smart Infrastructure and Applications. Springer, Cham, 2020. 115-133.
[2] Laña, Ibai, et al. "Road Traffic Forecasting Using NeuCube and Dynamic Evolving Spiking Neural Networks." International Symposium on Intelligent and Distributed Computing. Springer, Cham, 2018.
[3] Lee, Jooyoung, Ganghwa Kim, and Kitae Jang. "Estimation of Road Surface Condition during Summer Season Using Machine Learning." The Journal of The Korea Institute of Intelligent Transport Systems 17.6 (2018): 121-132.
[4] Ali Khalil, Mohamad, Khaled Hamad, and Abdallah Shanableh. "Developing Machine Learning Models to Predict Roadway Traffic Noise: An Opportunity to Escape Conventional Techniques." Transportation Research Record (2019): 0361198119838514.
[5] Elfar, Amr, Alireza Talebpour, and Hani S. Mahmassani. "Machine learning approach to short-term traffic congestion prediction in a connected environment." Transportation Research Record 2672.45 (2018): 185-195.
[6] Surya, S., and N. Rakesh. "Traffic Congestion Prediction and Intelligent Signalling Based on Markov Decision Process and Reinforcement Learning." Proceedings of International Conference on Cognition and Recognition. Springer, Singapore, 2018.
[7] Kong, Xiangjie, et al. "LoTAD: Long-term traffic anomaly detection based on crowdsourced bus trajectory data." World Wide Web 21.3 (2018): 825-847.
[8] Kumar, Priyan Malarvizhi, et al. "Ant colony optimization algorithm with Internet of Vehicles for intelligent traffic control system." Computer Networks 144 (2018): 154- 162.
[9] Lin, Lu, et al. "Road traffic speed prediction: a probabilistic model fusing multi-source data." IEEE Transactions on Knowledge and Data Engineering 30.7 (2018): 1310- 1323.
[10] Moriya, Koichi, Shin Matsushima, and Kenji Yamanishi. "Traffic risk mining from heterogeneous road statistics." IEEE Transactions on Intelligent Transportation Systems 99 (2018): 1-14.
[11] Zhang, Zhenhua, et al. "A deep learning approach for detecting traffic accidents from social media data." Transportation research part C: emerging technologies 86 (2018): 580-596.
[12] Poornappriya, T. S., and M. Durairaj. "High relevancy low redundancy vague set-based feature selection method for telecom dataset." Journal of Intelligent & Fuzzy Systems Preprint: 1-18
[13] Durairaj, M., and T. S. Poornappriya. "Choosing a spectacular Feature Selection technique for telecommunication industry using fuzzy TOPSIS MCDM." International Journal of Engineering & Technology 7.4 (2018): 5856- 5861.
[14] Durairaj, M., and T. S. Poornappriya. "Importance of MapReduce for Big Data Applications: A Survey." Asian Journal of Computer Science and Technology 7.1 (2018): 112-118.
[15] Durairaj, M., and T. Poornappriya. "A Review on Big Data Analytics Tools for Telecommunication Industry." International Journal of Circuit Theory and Applications 9 (2016): 185-193.
[16] Kumar, Akshi, and Renu Khorwal. "Firefly algorithm for feature selection in sentiment analysis." Computational Intelligence in Data Mining. Springer, Singapore, 2017. 693- 703.
[17] Sawhney, Ramit, Puneet Mathur, and Ravi Shankar. "A firefly algorithm based wrapper-penalty feature selection method for cancer diagnosis." International Conference on Computational Science and Its Applications. Springer, Cham, 2018.
[18] Bauer, Silke, and Marcel Klaassen. "Mechanistic models of animal migration behaviour–their diversity, structure and use." Journal of Animal Ecology 82.3 (2013): 498-508.
[19] Dingle, Hugh. "Animal migration: is there a common migratory syndrome?." Journal of Ornithology 147.2 (2006): 212-220.