IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received June 6, 2019, accepted June 13, 2019, date of publication July 2, 2019, date of current version August 7, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2926444

Integration of Ensemble and Evolutionary Machine Learning Algorithms for Monitoring Diver Behavior Using Physiological Signals

AFSANEH KOOHESTANI[1], MOLOUD ABDAR[2], ABBAS KHOSRAVI[1], (Member, IEEE), SAEID NAHAVANDI[1], (Senior Member, IEEE), AND MAHEREH KOOHESTANI[3]

[1]Institute for Intelligent Systems, Research and Innovation (IISRI), Deakin University, Geelong, VIC 3216, Australia
[2]Department of Computer Science, University of Montreal in Quebec, Montreal, QC H2X 3Y7, Canada
[3]Department of Electrical Engineering, Islamic Azad University, Mashhad 91735-413, Iran

Corresponding author: Afsaneh Koohestani (akoohest@deakin.edu.au)

ABSTRACT The level of consciousness and the concentration of drivers while driving play a vital role for reducing the number of accidents. In recent decade, in-vehicle infotainment (IVI) [or in-car entertainment (ICE)] is one of the main reasons that lead to degradation of drivers performance and losing awareness. However, the impacts of some other reasons, such as drowsiness and driving fatigue, are entirely important as well. Hence, early detection of such performance degradation using different methods is a very hot research domain. To this end, the data set is collected using two different simulated driving scenarios: normal and loaded drive (17 elderly and 51 young/35 male and 33 female). This paper, therefore, concentrates on driving performance analysis using various machine learning techniques. The optimization part of the proposed methodology has two main steps. In the first step, the performances of the K-nearest neighbors (KNN), support vector machine (SVM), and naive Bayes (NB) algorithms are improved using bagging, boosting, and voting ensemble learning techniques. Afterward, four well-known evolutionary optimization algorithms [the ant lion optimizer (ALO), whale optimization algorithm (WOA), particle swarm optimization (PSO), and grey wolf optimizer (GWO)] are applied to the system for optimizing the parameters and as a result enhance the performance of whole system. The GWO-voting approach has the best performance compared to other hybrid methods with the accuracy of 97.50%. The obtained outcomes showed that the proposed system can remarkably raise the performance of the classical algorithms used.

INDEX TERMS Diver behavior, machine learning, ensemble learning, evolutionary optimization algorithms (EOAs), physiological signals.

I. INTRODUCTION

Machine Learning (ML) is a reliable method searching in the large dataset to find unknown patterns [1]. ML has wide applications in various fields such as business, health care system, security and so on [2], [3]. Monitoring of diver behaviors is one of the recent applications of ML techniques [4]. Indeed, there are different techniques in ML which are applied to generate a model capturing the diversity between in behavior when drivers are aware with when they are distracted. There are different learning algorithms generating a single classifier such as support vector machine (support vector machine(SVM), K-nearest neighbors (KNN) and Naive Bayes (NB). These algorithms can be used as a prediction approach for new samples (unseen data). It is worth mentioning that, the performance of these classifiers may be influenced by the structure of some models in terms of initial model parameter settings. Even though choosing the single classifier by having the best performance is a valuable option, it may not work very well for all new samples. In most of the cases, the distribution of the new samples may be changed based on their applications. Recently, Recently, ensemble learning techniques have been proposed to combine several single ML methods to make powerful model in which has outstanding performance [5], [6]. In ensemble classifier, the predictions of single classifiers combine to generate the final prediction. One disadvantage of individual classifier is that some important information may be discarded while

The associate editor coordinating the review of this manuscript and approving it for publication was Mehul S. Raval.

robust solution is provided in an ensemble classifier. In this study, therefore, first three well-known ML single algorithms (SVM, KNN and NB) are applied to the diver Physiological signals.

The main objective of the current research is to find out whether drivers performance degradation can be predicted by an embedded system comprising of driver and vehicle information. The suggested proposed model, based on an ensemble of different classifiers and four evolutionary optimization algorithms (EOAs), can monitor drivers performance. After that, these three algorithms are combined using three ensemble learning techniques: Bagging, Boosting and Voting. Finally, the parameters of ensemble learning techniques are optimized using four EOAs. In this step, each single classifier obtained different weights using EOAs. The performance of each step is calculated and finally compared. The final results show that integration of ensemble and evolutionary ML algorithms can reach very high accuracy. The obtained outcomes indicated that combination of ensemble learning, and EOAs can significantly improve the performance of individual classifiers. To the best our knowledge, this is the first study in the domain of driver monitoring by using hybrid of ensemble learning and EOAs. Such a finding/insight can contribute to the development of a road map towards appropriate hardware and software for a reliable monitoring system. Eventually this will result in affordable monitoring systems which could be easily installed on new and existing vehicles.

The rest of this paper is organized as follows. Section III describes experiments conducted for collection of the data. It also provides some information about recorded physiological signals. Different optimization algorithms are introduced in Section IV. In section V, different machine learning algorithms and proposed method are briefly introduced for drivers' performance evaluation and findings are discussed and compared with other previous work in VII. Section VIII concludes the paper and summarizes the findings.

II. LITERATURE REVIEW

In this section, some of the recent studies in the domain of ensemble learning, EOAs and ML applications for driver behavior monitoring will be briefly reviewed. To do so, the application of ensemble and EOAs is discussed first. Then, the recent studies in the domain of driver behavior analysis using ML algorithms is presented.

A. ENSEMBLE APPLICATIONS

Ensemble classifiers have wide applications in different fields such as remote sensing, medical, transportation and so on [7]. Studies in this are is not always easy since they need sufficient range of data to avoid complex model caused by oversampling and simple model caused by under_sampling. Ensemble method is used in remote sensing because of several reasons: Firstly, a large amount of data from spaces is collected by satellite. Secondly, some data have poor label or do not have label. Finally, data includes a large number of feature and output such as (agriculture, water, forest) [8]. In the analysis of remote sensing, random forest (RF) is used since it selects the best features and then discards unused features. Majority voting for agricultural land shows that ensemble of three classifiers such as neural network (NN), different types of NN such as multi layer perceptron (MLP) probabilistic neural network (PNN), Gaussian classifier and KNN could worked much better than single classifiers [9]. An other application of ensemble method is in person recognition [10]. This method is used to identify the characteristic of person and behavior. The amplitude and frequencies of signals collecting from speech, face and human behavior recognition should be captured to extract new features. However, the combination of multi features into the one feature is not easy since they have multiple scales. Ensemble method includes individual recognizers for each modality and guarantee the performance of the model because they combine at the decision level where the scales would be the same. An other limitation for using single classifier in person recognition is collection of good data [11]. Almost, measurement instruments introduce noise and decrease the performance of individual classifier. The ensemble method is used to detect of speech and face and driver's behavior in the vehicle, the results are shown that the ensemble of several classifiers outperformed the single classifiers. In medical application, ensemble method is used due to the limitation of training and testing samples, imbalance dataset and different misclassification costs. In pharmaceutical molecule classification, MRI images and EEG signals showed that the ensemble of individual classifiers is more informative than single ones [12].

B. EVOLUTIONARY APPLICATIONS

Recently, many papers apply nature-inspired stochastic optimization to improve the performance of their proposed models. These techniques are inspired by the laws of the natural evolution or the social behavior of a group of animals. The optimization process with random variables is used to generate random solution. The main important point is that the best individuals are combined together in each generation to generate the next generation. This allows the population to be optimized over the course of generations. The process is terminated when the best solution is founded. They have a wide range of applications in different fields due to their performances. It is said that there is not unique optimization technique used to solve optimization problems. Some common optimization methods are known such as particle swarm optimization (PSO) [13], ant lion optimization (ALO) [14], grey wolf optimization (GWO) [15], wale optimization algorithm (WOA) [16]. One benefits of using PSO is to preserve information over the next iterations while other EOAs eliminate their search space information after generating the new generation. The search process is divided into two stages are known as exploration and exploitation [17]. The optimizer uses operators to globally explore the search space: in this stage, movements are chose randomly. After exploration phase, the exploitation is investigating the regions of the search space. Exploitation depends on the local search capability in the promising regions of design space found in the exploration phase. One of the challenging task in the estimation of EOAs is to find a proper balance between exploration and exploitation [18]. There are many applications of EOAs in different industries, such as PID controller in DC motors. [19], analysis the surface wave [20], stabilize the wide- area power system [21], and solve load frequency control problems [22]. In addition, biogeography based optimization (BBO) is used for solving the feature selection problem [23]. Another application of EOAs is in energy section such as electrical power, solar energy which are discussed in [24], [25]. In medical industry, Genetic Algorithm (GA) is applied to detect [26] heart disease. Researchers in [27] applied the heuristic optimization algorithm to control the driverless vehicle by using multi agent system. A novel grey wolf optimizer which is introduced in [28] which is the upgrade version of GWO. In this algorithm, some modifications have been done to enhance the convergence of the GWO model. It includes four categories. The first part is modification of some control parameters while the second one is hybridization with other search methods. The position updating is in the third part. The last part is introduction to the new operators. Researchers in [29] developed a new version of GWO which is called alpha-guided grey wolf optimizer (AGGWO). To do so, they modified the alpha since it plays remarkable role because alpha is more responsible for making decisions compare to other control parameters such as beta, delta and gamma. Another meta heuristic algorithm is called a modified sine cosine algorithm (m-SCA) which is introduced by Gupta and Deep [30]. To solve the optimization problem, characteristics of sine and cosine trigonometric are used. The classical SCA suffers from the slow changes in local optimum point. Therefore, this study solves the problem by generating the opposite population and after that, a self-adaptive direction is integrated into the search equation. A novel random walk grey wolf optimizer (RWGWO) is developed by [31]. This algorithm is the upgrade version of swarm intelligence to solve the continuous optimization problems. Gupta and Deep concentrated on improvement among leaders to enhance the convergence. Thus, they introduced the random walk which explores the space and permits omega wolves to update their position. In the field of power system, [24], submulti optimizer is introduced to achieve better efficiency of stability and convergence when there is a mismatch between manufacturers tolerances of PV cell characteristics and different shading luminosity. A memetic salp swarm algorithm (MSSA) is introduced by [25] as the upgraded version of original salp swarm algorithm. The main application of this algorithm is used for enhancing the output of the power systems as it influenced by partial shading condition and fast-time weather conditions. Therefore, MSSA applies multiple independent slap chains to the model to have better convergence and maximum stability. Furthermore, Yang et al [32], developed a grouped grey wolf optimizer (GGWO) to increase the output of wind turbine. The main difference between the GGWO and GWO is that GGWO increases the speed of convergence compared to GWO.

C. DRIVING MONITORING

Driving is one of challenging tasks that may provide some challenges for passengers travelling with their own vehicles. According to the reports, lane departure due to degradation in driver performance is one of the most common vehicle accidents worldwide [33] and took many lives [34].

Apart from the benefits of advanced driver assistant systems (ADAS), it may increases some challenges for drivers since they do not have much attention/control on their vehicles. As a result, it leads to increase the number of road accident [35]. To prevent this issue, a reliable method should be developed to monitor driver's behavior. Some ML methods have been studied to evaluate driver performance [36].

Lane deviation as a response variable is shown by [37]–[39]. They introduced a trajectory model called video-based lane detection (VLD). This model alerts drivers after deviation from the central lane. The limitation of these studies are 2-folds. Firstly, this model cannot identify the lane departure with only one side lane boundary. Secondly, it makes a mistake when there is an arrow mark near the central lane. In our research, we solve this issue by introducing the standard lane boundaries from two sides of the road. This threshold separates normal from wavy driving if the driver goes over it. As showed in Table 1 various algorithms have been recently studied to monitor driver performance in normal or under cognitive conditions. In [40], detection reaction time (DRT) and brake reaction time (BRT) were considered to monitor driver's behavior. The issue of this study is to require additional apparent. Therefore, it increases the final cost of the vehicle. Physiological signals and vehicle information were fed into the HMM model and operation-triplet (OT) [41]. Multi-input and multi-modal methods are applied to increase the robustness of the model. However, these methods do not show significant performance if few samples are chosen. The same study, [42], recorded electrooculography (EOG) to detect fatigue through using autoregressive integrated moving average (ARIMA) method. The main barrier using EOG is sensitive to motion and noise. In another study, the vehicle information was chosen as an input data for the SVM and NN classifiers. The NN obtained low accuracy (69 %) compared to the SVM method [43]. Even though the SVM and NN classifiers achieved accuracies of 95 % and 96 % respectively in [44], they require a significant amount of data to train the model. EEG signal is studied by [45] to evaluate driver's performance. However, the main issue is that the EEG signal is less practical in the real scenario. In [46], the ECG, EMG, EDA and Respiratory were applied to LDA and fisher projection matrix (FPM) methods to recognize stress level of drivers. However, the accuracy was not high in showing variation. The last but not least, the result in [47] showed that the performance of SVM and NN were lower compared to Bagging method which

TABLE 1. Previous research methods for analysis drivers behavior.

Reference	Inputs	Methods	Results	Shortages
[40]	EEG, ERBP	DRT [1] BRT [2]	BRT and accuracy much better in DRT than BRT	Additional apparent in in the instrumented vehicle
[43]	Vehicle information	SVM, NN	Detect cognitive driver	Poor accuracy in classification
[46]	ECG, EMG EDA, RES	LDA [3], FPM [4] variations,	Detect stress level	Poor accuracy in showing variation
[45]	EEG, fNIRS	BPNN [5], NN	Detect fatigue	Sensitive to noise and motion
[39]	Vehicle information	VLD [6]	Detect lane deviation	Constraint in lane boundaries
[41]	Head movement, steering wheel angle, eye movement	HMM [7],OT [8]	Detect driver's activities	Fewer inputs reduced the performance of classifiers
[42]	EOG	ARIMA [9]	Detect fatigue	Sensitive to noise and motion
[50]	Gaze entropy	SEE [10],MCM [11]	Predict drowsy driver through the lane changes	Small sample size less realistic

[1] detection reaction time
[2] brake reaction time
[3] linear discriminative analysis
[4] fisher projection matrix
[5] back propagation neural network
[6] video lane detection
[7] hidden marco model
[8] operation triplet
[9] autoregressive integrated moving average
[10] shannon's entropy equation
[11] marco chain matrices

is confirmed in this research as well. SVM and adaptive neuro-fuzzy inference system (ANFIS) were chosen as classifiers to monitor driver's performance By Katsis *et al.* [48]. In this paper, EEG, ECG, EDA and RES were recorded from loaded drivers. The accuracies of them were 79 % and 76 % which is not as high as [45], [46], [49]. [50] applied gaze information to Shannon's Entropy Equation(SEE) and Marco Chain Matrices (MCM) to predict drowsy driver through the changes. The result is less realistic as they used small number of samples. This can be considered as a disadvantage for the proposed methodology of Shiferaw et al.

III. DATA ANALYSIS

A. DATA PREPARATION

The dataset used in this paper was collected by [51] and is publicly available for download through the Nature website (https://www.nature.com/articles/sdata2017110). In this dataset, 68 subjects completed the experiments which is done in a driving simulator. Participants were not reported any illness history. According to the data, one participant reporting motion sickness was excluded from experiments. The range age of ages of subjects was between 18-27 years old, and above 60 years old. The data collected is balanced based on the gender: 35 males and 33 females. The scenario of driving was taught to the participants. Each driver could be in the two main categories, distracted or normal based on their lane deviation.

As mentioned in [51], all the physiological and observational measurements were recorded unobtrusively. They were collected by using advanced thermal and visual cameras as well as wearable sensors. Accordingly, all subjects exhibited natural/normal behaviors during simulated driving experiments. During experiments, visual and thermal cameras were taking continuous images from the participants faces. Also, two wearable sensors recorded different physiological signals such as heart rate, palm EDA and breathing rate. Thermal facial imagery records were used to obtain perinasal perspiration. Visual facial camera was used to record gaze

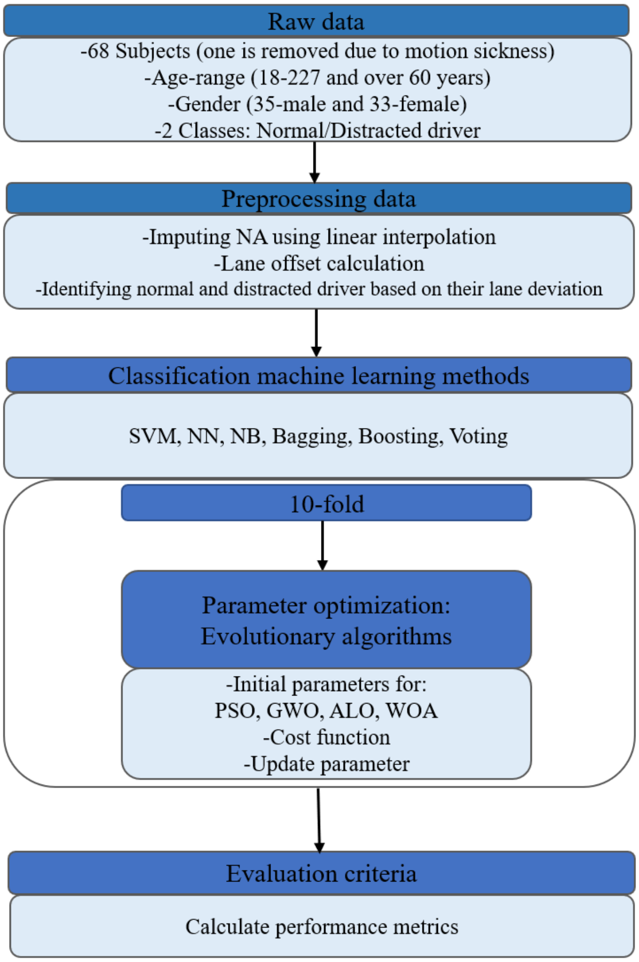


FIGURE 1. The general overview of the proposed methodology. Physiological data and vehicle information are selected as input features. The performance of the driver is determined by lane deviation. After preprocessing of input features, they are mixed with the response variable. Three classifiers and their ensembles before and after applying EOAs are applied into the model. In the final step, performance metric will be used to evaluate the performance of each classifier and their ensembles.

information and Pupil diameters. Vehicle information such as lane position, speed, brake, acceleration and steering were recorded by driving simulator. Fig. 1 shows the set up used for experiments and data collection. Some useful information about collected is provided in Table 2.

Experiments are two-fold:

- Normal scenario: drivers drove normally. They were not under additional stressor.
- Loaded scenario: drivers did text back words while driving.

B. ANALYSIS OF DRIVER'S BEHAVIOUR

Here, both lane offset and lane position were selected to show the impact of distraction on driver's behavior. First normal drivers are shown regarding to the performance while driving without any secondary activities. If drivers drive more than 0.2% of margin, they are excluded. Afterwards, the lane offset

TABLE 2. The feature of dataset used in our study.

Input features	min	max	mean	std
palm.EDA	29.1097	$5.858e^5$	$1.2736e^4$	$7.3388e^4$
heart rate	0	151	74.7169	24.5733
breathing rate	10.6000	30.4000	20.3046	3.2883
perinasal perspiration	0.0037	0.0188	0.0067	0.0025
speed	10.8382	86.7498	6.5083	0.0048
acceleration	-0.3345	37.1228	5.6095	3.7496
brake	0	267.5502	10.1103	0.18860
steering	-2.8587	6.2828	0.0024	42.5837
gaze.X.position	0	$1.8304e^3$	341.7007	408.9952
gaze.Y.position	0	$1.0795e^3$	317.4544	369.5284
left.pupil diameter	0	0.0079	0.0020	0.0024
right.pupil diameter	0	0.0091	0.0022	0.0025

was calculated for normal drivers. The width of lane and car are 3.65m and 1.85m, respectively. Accordingly, the boundary of drive is $\frac{3.65-1.85}{2} = 0.9$ m in each side. Accordingly, their driving may be impacted during the loaded drive if they deviate from the lane central line and passes 2/3rd of the boundary (0.6m). This is shown by:

$$\text{impacted} = \begin{cases} 1 & |\text{lane offset}| \geq 0.6 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

This conception is inspired by the results studied in [51]. The lane offsets show small positive or negative values in most of the experiments. Therefore, a threshold is defined to show deviation of the label moments when performance of driver has been degraded while driving under loaded conditions.

IV. OPTIMIZATION ALGORITHMS (EAS)

This section introduces several common EOAs.

A. PARTICLE SWARM OPTIMIZATION (PSO)

Particle Swarm Optimization (PSO) is one of the robust algorithm in evolutionary computation technique. This algorithm is introduced by Kennedy and Eberhart. As it is mentioned the inspiration source of it is based on the social behavior of bird flocking. It uses a number of particles solutions flying over the search space to find best solution. In other words, each particle take an account their solution as the best solutions they have taken. The PSO algorithm is mathematically defined as follows (see equations 2 and 3).

$$\begin{aligned} v_i^{t+1} &= wv_i^t + c_1 \times rand \times (p_{best}\_x_i^t) + c_2 \\ &\quad \times rand \times (g_{best}\_x_i^t) \end{aligned} \tag{2}$$

$$x_i^{t+1} = x_i^t + v_i^t \tag{3}$$

where the particle velocity for i_t particle at the iteration t, w is a weight function, and also c_j is a weighting factor. Moreover, rand stands for a randomly selected number between 0 and 1, x_i^t is the present position of the i_{th} particle at the t_{th} iteration, the p_{best} (the personal best position) is the best position of the particle i at the iteration t, and the g_{best} (the Global best) is the best solution found so far. far.

BASIC OPERATIONS OF PSO

The qualified numbers in each iteration called particles are combined to produce more qualified ones. Each of these particles would explore a potential solution to a problem. A decoding function explores a map for this genotype. Only the fittest individuals selected will be produced and pass their information to the next particles. Therefore, PSO picks up the best solution inspiring the natural behavior of members in a competitive environment. In this research the position updating is changing between "0" and "1" . The PSO algorithm is summarized by Algorithm 1.

Algorithm 1 Particle Swarm Optimization (PSO)

Set particle swarm optimization parameters
/*
Size of population = X_i, (i=1,...,n)
Initialize the particle's position: lower bound< $ParticleBest.Position_i$ <upper bound
Initialize the particle's best known position to its initial position: Particle.position ← xi
Initialize the particle's velocity
*/
Definition of fitness function
/*
$functionJ = computeCost(X, y, w)$
$J = 0$
$prediction = X * w$
$sqrError = (prediction - y)^2$
$m = length(y);$ $J = 1/(2 * m) * sum(sqrError)$
*/
/* X=input (ensemble training)
y=Response variable
w= Weights for three types of classifiers */
if $CostFunction(particle.Position_i) < CostFunction(GlobalBest)$
then
  Update particle's position
  Update particle's velocity
/* $Velocity_{i+1}$= w*$Velocity_i$+c1*rand(VarSize)*($ParticleBest.Position_i$ $-particle.Position_i$) +c2*rand(VarSize).*($GlobalBest.Position_i$ $-Particle.Position_i$); */
/* $Particle.Position_{i+1}$= $Particle.Position_i$+ $Velocity_{i+1}$ */
else
  Stop updating

Calculate weights for all classifiers
/* w_{KNN}, w_{SVM}, w_{NB} */

B. ANT LION OPTIMIZER (ALO)

Antlions belong to the Myrmeleontidae family and net-winged insects. The origin of their names coming from the unique hunting behavior. They try to adapt themselves to be survival with digging out traps. The larger traps show that they are more hungrier. The Ant Lion Optimizer (ALO) algorithm imitates interaction between antlions and ants. In this interaction, antlions hunt ants when moving over the search space. In the next phase, the behavior of antlions and their prey in nature is proposed by mathematical definition. This optimizer tends to find optimal solutions for problems when applying random solutions [14], [52].

1) BASIC OPERATIONS OF ALO

Two populations are introduced in the ALO algorithm: One population is ant and the other one is antlion. In the next step, the values of these populations should be changed to estimate global optimum value. Firstly, the ant population is randomly selected as it is the main search agents in this optimizer. Secondly, cost function is used to evaluate fitness value in each iteration. Thirdly, ants randomly walk around the antlions when moving over the search region. Finally, the number of antlions is not considered. It is imagined the position of antlions and ant are the same in the first iteration. Each antlion belongs to each antand updates the position if the ants become fitter. The distance among ants are affected by an elite antlion. Thus, if each antlion becomes better it will be replaced with prime one. These steps continuous until it converges. Then, the fitness value and position of the elite antlion are selected as the best estimation for the global optimum [14], [53].

$$X(t) = [0, cumsum(2r(t_1) - 1), \ldots, cumsum(2r(t_n) - 1)] \quad (4)$$

where t and n show the random walk and maximum number of iteration and cumsum calculates the cumulative sum (see equation 4).

$$\text{r(t)} = \begin{cases} 0 & \text{if rand} \geq 0.5 \\ 1 & \text{otherwise} \end{cases} \quad (5)$$

Random walk is normalized by the following equation (see equation 5), where c_i^t is the minimum of i-th variable at t-th iteration, d_i^t indicates the maximum of i_{th} variable at t_{th} iteration, a_i^t is the minimum of random walk of i_{th} variable, and b_i^t is the maximum of random walk in i_{th} variable (see equation 6).

$$X_i^t = \frac{(x_i^t - a_i) \times (d_i^t - c_i)}{b_i - a_i} \quad (6)$$

Equations 7 and 8 show that ALO simulates the entrapment of ants in antlions pits where $Antlion_j^t$ shows the position of the selected j_{th} antlion at t_{th} iteration.

$$c_i^t = Antlion_j^t + c^t \quad (7)$$
$$d_i^t = Antlion_j^t + d^t \quad (8)$$

In the ALO, roulette wheel is used to choose fitter antlion to find fitness value. Therefore, the margin for random walk should be decreased as follows: $I = 1 + 10^w \frac{t}{T}$ where t is the current iteration, T is the maximum number of iterations, and

w is defined based on the current iteration adjust the accuracy (see equations 9 and 10)

$$c^t = \frac{c^t}{I} \tag{9}$$

$$d^t = \frac{d^t}{I} \tag{10}$$

In equation 11, it is shown that the last operator is to find the fittest antlion and store it. Ant_i^t shows the position of i_{th} ant at t_{th} iteration, R_A^t shows the random walk around the fitter antlion, and R_E^t shows random walk around the elite at t_{th} iteration.

$$Ant_i^t = \frac{R_A^t + R_E^t}{2} \tag{11}$$

The ALO algorithm is summarized by Algorithm 2.

Algorithm 2 Ant Lion Optimizer (ALO)

```
Set Ant Lion Optimizer parameters
/*
Size of populationfor ants and antlions
Select a random antlion
Select the elite using Roulette wheel
Create a random walk and normalize
                                                        */
While t< max number of iterations
Definition of fitness function
/*
functionJ = computeCost(X, y, w)
J = 0
prediction = X * w
sqrError = (prediction − y)^2
m = length(y);  J = 1/(2 * m) * sum(sqrError)
                                                        */
/* X=input(ensemble training)
y=Response variable
w= Weights for three types of classifiers               */
if CostFunction(Ant_i^t) < CostFunction(antlion_j^t)
then
    Update the position of ant
else
    Stop updating

Calculate weights for all classifiers
/* W_KNN, W_SVM, W_NB                                   */
/* t=t+1
                                                        */
End While
```

C. GREY WOLF OPTIMIZER (GWO)

The grey wolf optimizer (GWO) is introduced by [15]. The social behavior of grey wolves and the technique of their hunting's are considered to solve optimization problems. They are known as apex predators since they are placed on the top of the food chain. The origin of coming from Canidae family. Their average size is around 5-12. Therefore, the mathematical equation is defined to describe the social hierarchy of wolves. The fittest solutions are known alpha, beta, and omega. Other solutions are considered as delta [54].

1) BASIC OPERATIONS OF GWO

The GWO are categorized into the four types of wolves. In each group, they have a leader (first type of wolf) which is responsible for making a decision and other wolves follow his/her leaders. This leader is called alpha in a pack. Only alpha is allowed to mate in a pack. However, the alpha may not be strongest in the member, it can be the fitter one to manage other wolves in a pack. This is show that the structure and order are important factor in alpha's decision. The second type is called beta helping the alpha in making decision and other wolves in a pack. Moreover, he/she can be replaced by one of the alphas after passing away or getting older. The third type is called omega. This omega type devoted to wolves which are scapegoat in a pack. The pack may be faced with a problem such as fighting if it loses the omega. Other wolves belong to the delta type. They are responsible for watching the boundaries and warn if they feel any dangers, protect and increase the safety in a pack, providing food for a pack and caring the ill or weak wolves [15].

$$\overrightarrow{D_\alpha} = |\overrightarrow{C_1}.\overrightarrow{X_\alpha} - \overrightarrow{X(t)}| \tag{12}$$

$$\overrightarrow{D_\beta} = |\overrightarrow{C_2}.\overrightarrow{X_\beta} - \overrightarrow{X(t)}| \tag{13}$$

$$\overrightarrow{D_\gamma} = |\overrightarrow{C_3}.\overrightarrow{X_\gamma} - \overrightarrow{X(t)}| \tag{14}$$

α, β and γ are shown the distance from the prey in equations (12, 13 and 14) where t is the current iteration, $\overrightarrow{A}$ and $\overrightarrow{C}$ are coefficient vectors, $\overrightarrow{X(p)}$ (p $=$ α (equation(15)), β (equation(16)) and γ (equation(17))) is the position vector of the prey, and $\overrightarrow{X}$ indicates the position vector of a grey wolf (see equations 18, 19 and 20).

$$\overrightarrow{X_1} = |\overrightarrow{X_\alpha} - \overrightarrow{A_1}.(\overrightarrow{D_\alpha})| \tag{15}$$

$$\overrightarrow{X_2} = |\overrightarrow{X_\beta} - \overrightarrow{A_2}.(\overrightarrow{D_\beta})| \tag{16}$$

$$\overrightarrow{X_3} = |\overrightarrow{X_\gamma} - \overrightarrow{A_3}.(\overrightarrow{D_\gamma})| \tag{17}$$

$$\overrightarrow{X(t+1)} = \frac{\overrightarrow{X_1} + \overrightarrow{X_2} + \overrightarrow{X_3}}{3} \tag{18}$$

$$\overrightarrow{A} = 2\overrightarrow{a}.\overrightarrow{r_1} - \overrightarrow{a} \tag{19}$$

$$\overrightarrow{C} = 2\overrightarrow{r_2} \tag{20}$$

The value of $\overrightarrow{a}$ and $\overrightarrow{C}$ are random value between [0,2] $\overrightarrow{r_1}$ and $\overrightarrow{r_2}$ are random vectors in [0,1]. The GWO algorithm is summarized by Algorithm 3.

D. WHALE OPTIMIZATION ALGORITHM (WOA)

Whales are one of the smart mammal animals [16]. They have some common cells in their brain which is like the humans. These play key roles in judgment, emotions and social behaviors of whales. Whale have twice number of these cells in their brains compared to humans. This feature makes them smartness. However, the level of their judgments, thinking and emotions are lower than humans. They can live

Algorithm 3 Grey Wolf Optimizer (GWO)

Set Grey Wolf Optimizer parameters
/* Size of population = X_i, (i=1,...,n)
The best agent = X_α, X_β, X_γ
A,C = Coefficient vectors
a= Decreasing value between [0,2]
*/
While t< max number of iterations
if $|A| < 1$
then
  Update the position of current search agent
else
  Stop updating
Definition of fitness function
/*
$functionJ = computeCost(X, y, w)$
$J = 0$
$prediction = X * w$
$sqrError = (prediction - y)^2$
$m = length(y)$; $J = 1/(2*m) * sum(sqrError)$
*/
/* X=input (ensemble training)
y=Response variable
w= Weights for three types of classifiers */
if *CostFunction(previous search agent) < CostFunction(current search agent)*
then
  Update X_α, X_β, X_γ
else
  Stop updating

Calculate weights for all classifiers
/* W_{KNN}, W_{SVM}, W_{NB} */
/* t=t+1
*/
End While

alone or in a group. They are almost in a group. Humpback whales is the biggest in the whales group. They hunt small fish based on their hunting which can be used for solving the optimization problems [16].

1) BASIC OPERATIONS OF WOA

Humpbacks prefer to hunt krill and small fish herds. They have special hunting method which is called bubble-net feeding method [55]. They hunt krill or small fishes are closer to the surface through distinguishing bubbles along a circle. They are two manoeuvres cooperated with bubble and called upward-spirals and double-loops. The position of the optimal target is not known and the current position is considered as an optimal solution. Afters searching, the humpbacks update their positions. In the next step, they attack to their prey with two methods, shrinking encircling mechanism and spiral updating position. In the former one, the new position is

TABLE 3. The parameters of evolutionary algorithms applied in this study.

Optimization algorithm	PSO	ALO	GWO	WOA
Number of agent	30	30	30	30
Number of iteration	500	500	500	500
Upper bound	10	10	10	10
Lower bound	-10	-10	-10	-10
a	-	-	[2 to 0]	[2 to 0]
r	-	[0,1]	[0,1]	[0,1]
v_{max}	6	-	-	-
c_1	2	-	-	-
c_2	2	-	-	-

defined between the original position and the current position of the best agent. In the later one, the distance between the position of whale and location of prey is calculated. Finally, to search for prey, the position of a search agent will be randomly updated instead of the best search agent found so far.

$$\overrightarrow{D'} = |\overrightarrow{C}.\overrightarrow{X} - \overrightarrow{X(t)}| \tag{21}$$

$$\overrightarrow{X(t+1)} = \begin{cases} \overrightarrow{D'}.\overrightarrow{e^{bl}}cos(2\pi.l) + \overrightarrow{X^*(t)} & \text{if p} \geq 0.5 \\ \overrightarrow{X(t+1)} = |\overrightarrow{X^*(t)} - \overrightarrow{A}.(\overrightarrow{D})| & \text{otherwise} \end{cases} \tag{22}$$

where t shows the current iteration, $\overrightarrow{D'}$ shows the distance of the i th whale to the prey, b is a constant, l is a random number in [-1,1]

$\overrightarrow{A}$ and $\overrightarrow{C}$ are coefficient vectors, $\overrightarrow{X^*(t)}$ is the position vector of the best solution obtained should be updated if a better solution is founded. The vectors A and C are calculated as follows:

$$\overrightarrow{A} = 2\overrightarrow{a}.\overrightarrow{r} - \overrightarrow{a} \tag{23}$$

$$\overrightarrow{C} = 2\overrightarrow{r} \tag{24}$$

The value of $\overrightarrow{a}$ is decreasing value between [0,2], $\overrightarrow{r}$ and p are random vector and numbers in [0,1]. The WOA algorithm is summarized by Algorithm 4. All important information for design these EOAs is summarized by Table 3.

V. CLASSIFICATION ALGORITHMS

In this section, the data mining methods used to determine whether the driver's performance is degraded or not will be briefly explained. Sections 5.1 - 5.3, briefly introduce three classification algorithms. All these methods have been selected due to their promising performance as reported in the literature. The performance of these algorithms is well demonstrated through several forecasting and classification competitions [56], [57]. The new hybrid ensemble is proposed in Section 4.12 to derive precise evaluation.

A. K-NEAREST NEIGHBOURS (KNN)

This classifier finds the K variables in the closest training dataset. The distance between training and testing data is defined based on the weight to the contributions of the neighbors. The response variable and input features are needed for

Algorithm 4 Whale Optimization Algorithm (WOA)

```
Set Whale Optimization algorithm parameters
/*
Size of population = X_i, (i=1,...,n)
Search the best agent = X*
A,C = Coefficient vectors
a= Decreasing value between [0,2]
l= Random number[-1,1]
                                                        */
While t< max number of iterations
if p<0.5 and |A| < 1
then
    Update the position of current search agent
else
    |A| ≥ 1 and p ≥ 0.5; Choose the random search
    agent and update the position of the current search
Definition of fitness function
/*
functionJ = computeCost(X, y, w)
J = 0
prediction = X * w
sqrError = (prediction − y)^2
m = length(y);  J = 1/(2 * m) * sum(sqrError)
                                                        */
/* X=input(ensemble training)
y=Response variable
w= Weights for three types of classifiers               */
if CostFunction(previous search agent) <
CostFunction(current search agent)
then
    Update X*
else
    Stop updating

Calculate weights for all classifiers
/* W_KNN, W_SVM, W_NB                                   */
/* t=t+1
                                                        */
End While
```

this algorithm. The input variables K determines the quantity number of neighbors which are necessary. Finally, the output is determined by the greatest common K nearest training to testing [58]

B. SUPPORT VECTOR MACHINE (SVM)

Support vector machines (SVM) is a supervised learning algorithm which is used for either linear or non-linear classification challenges. In this algorithm, a nonlinear matching method is applied to the original dataset. This algorithm transforms dataset in to the higher dimension. The higher dimension is called hyperplane and used for decision boundary. One benefits of using SVM is to calculate an optimal decision boundary to shuffle data into different classes with high relevant features [59].

C. NAIVE BAYES (NB)

The Naive Bayes (NB) algorithm is one of the powerful classifiers working based on Bayes' theorem. The probabilistic information is used to learn, and represent, to reduce the unambiguously of the statistical components in this method. This algorithm has wide applications in different field as it works independently. The algorithm has good performance compared to other individual classifiers, such as decision trees and neural network [59].

D. VOTING-ENSEMBLE

Voting gives an opportunity to combine several individual classification algorithms. In general, Voting ensembles learning technique includes un-weighted and weighted terms. In un-weighted term, it works based on the Min and Max,voting, average and product of probabilities. In the weighted term includes simple weighted voting, the weight for individual classifier is calculated based on the prediction of classifier with its training dataset. Rescaled weighted voting, zero weight is devoted to the classifier that it has less correct performance. Therefore, classifiers with lower than threshold are excluded form the ensemble. The best worst weighted voting, the threshold for this classifier is between [0,1]. The highest weight is assigned to the classifier with high performance and lowest weight belongs to the classifier with the lowest performance. Weighted voting is widely used in different applications [60]. In this method, the weight of output of classifiers are not the same. the higher weight values show the better performance of classifiers [5]. Therefore, it is very important to find appropriate weights for all classifiers. Weighting problem can be solved by EOAs such as PSO, WOA, GWO and ALO.

E. BAGGING-ENSEMBLE

In the Bagging algorithm, new datasets are produced by taking bootstrap samples from the original dataset. This algorithm trains classifier on each sample. Finally, it is combined by using a majority voting with new sampled datasets and combined by using a majority voting combination rule. Even though Bagging is very simple method, it is a strong method for a some datasets. In addition, it is powerful for unstable learning algorithms. The output of this classifier is the combination of posterior probability for each classifier. The famous Bagging classifier is called Random forest. It works based on the structure of decision trees on the bootstrap samples. This diversifies how trees are grown, so each captures some specific patterns from the resampled set. The final outcome of the model is generated by an aggregation mechanism such as voting or averaging. It is already known that a collection of parallel decision trees outperforms each individual tree [61].

F. BOOSTING-ENSEMBLE

Ensemble method implements multiple learning algorithms to improve the performance of final prediction. This method supervises aggregated models and learns how to best combine

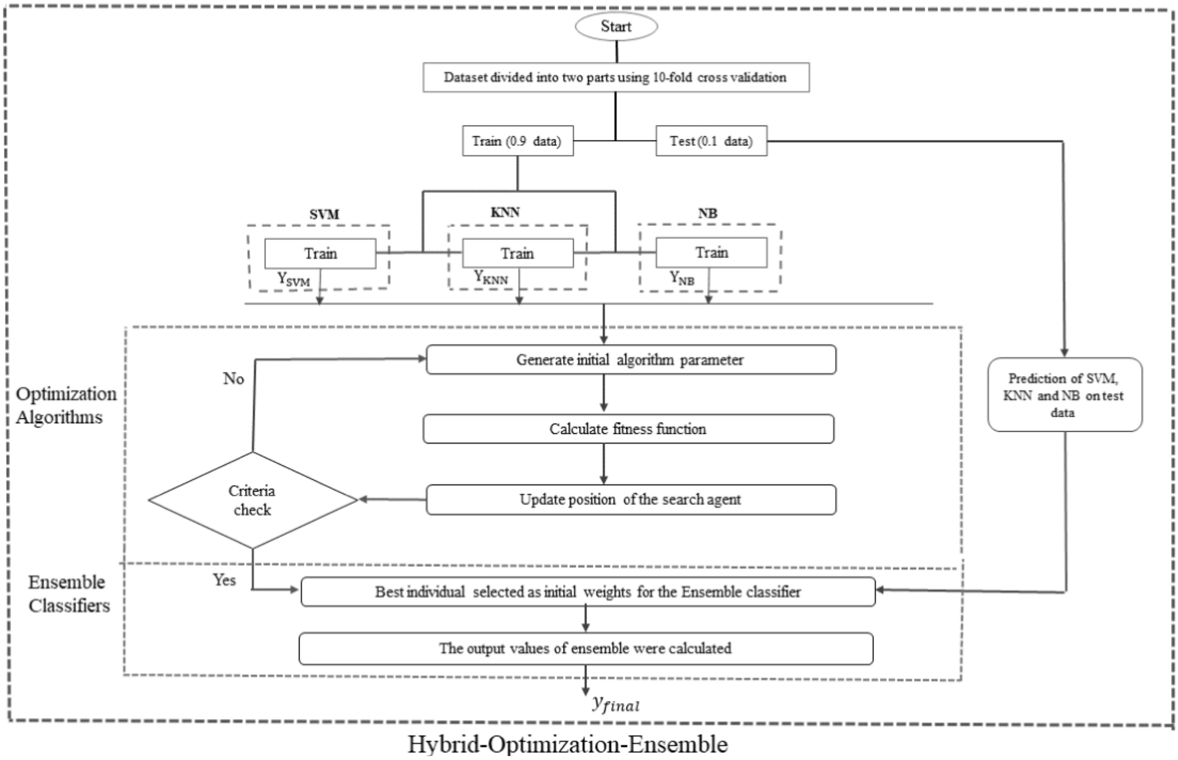


FIGURE 2. The internal architecture of the proposed methodology. Dataset is divided into two parts (0.9 data is used for training and 0.1 for testing). Three classifiers such as SVM, KNN, NB are used to evaluate the performance of the driver's data. Their outputs are forwarded to the for EOAs to obtain weights for different classifiers. These weights are fed in to the Bagging, Boosting, and Voting ensemble learning techniques. The final result is used for evaluation of driver's performance.

the predictions of the primary models. After training several learning algorithms, it combines the prediction of various learning algorithms. All algorithms such as SVM, KNN and NB are trained. The prediction of these models are combined and used for making the final prediction. Boosting ensemble uses the logic that the next predictors is trained from the mistakes of previous steps. As a result, the observation appears with unequal probability of the model with the highest error. In other words, the observation is selected based on error. The predictors can be selected from some classification algorithms. The process of this method is short as it works with the mistakes of previous predictors [6].

G. PROPOSED METHOD

According to Fig.2 and Algorithm 5, the data will be divided into two major parts (using 10-fold cross validation): train and test data. in the second level, we applied KNN, SVM and NB on the train data. Then, the prediction of these algorithms is given to the EOAs introduced in the previous section for calculating the weights of these algorithms.

the output of these optimizers will be applied to the ensemble methods. The first step in EOAs is definition of population size. Secondly, other required parameters should be randomly initialized. Afterwards, the fitness function is computed based on the RMSE model. In this research, we tried to find the minimum RMSE. After finding the minimum value, the search agent stops. The outputs are the initial weights for proposed ensemble methods. Finally, the ensemble method predicts the class of each test data using obtained weights by EOAs and prediction of KNN, SVM and NB. The fitness function tries to identify the optimal global point.

VI. RESULTS

The proposed methodology is conducted in the two main experiments. In the first experiment, three simple machine learning algorithms are applied. Afterward, the performance of those algorithms is improved using three ensemble learning techniques. While in the second step, four well-known EOAs are applied to the previous step to achieve better outcomes. The obtained results of each step are presented in following sub-sections, respectively.

A. FIRST EXPERIMENT

In this phase of study, the proposed methodology will be applied to the dataset. To this end, three well-known machine learning algorithms are applied: KNN, SVM and NB algorithms. The obtained outcomes are presented in Figure 3 and Table 4. According to these information, it is obvious that NB has batter performance compared to other two algorithms. Moreover, it can be seen that SVM has very close performance to NB, whereas KNN has weaker performance than other two algorithms. To enhance the prediction performance,

Algorithm 5 Proposed Algorithm (Integration of Ensemble and Evolutionary Machine Learning Algorithms)

```
Input: Matrix with raw input-data
/* physiological and vehicle information */
Output: Response variable
/* lane deviation */
if lane deviation >2/3 of the margine (the center line)
then
    1 /* driver is distracted */
else
    0 /* Driver is not distracted */
Data: Training and Testing dataset
Introduce cross-validation for K=10
Introduce three types of classifiers
/* KNN,SVM and NB */
Evaluation coefficient and confusion matrix for actual
and prediction data
/* Accuracy, ROC, Recall,Precision, F-score */
Create ensemble
/* data is divided into ensemble trianing and
   ensemble testing set */
Stack three classifiers by Bagging, Boosting, and Voting
method
/* apply ensemble trainig set to the Bagging,
   Boosting and Voting classifiers */
Evaluation coefficient and confusion matrix for actual
and prediction ensemble data
/* Accuracy, ROC, Recall,Precision, F-score */
Apply weights obtained from Algorithm1 to the
ensembleTest data
if current predictor error> previous predictor
then
    continues learning
else
    stop learning
/* Decision condition by using Hybrid
   Optimization Ensemble (HOE) detection
   approach */
Evaluation coefficient and confusion matrix for actual
and prediction ensemble data
/* Accuracy, ROC, Recall,Precision, F-score */
```

TABLE 4. Comparison of the simple machine learning algorithms.

Metric	KNN	NB	SVM	Bagging	Boosting	Voting
Accuracy	0.8386	0.9089	0.8930	0.9351	0.9310	0.9400
Precision	0.8560	0.9435	0.9422	0.9838	0.9372	0.9838
Recall	0.7750	0.8836	0.8793	9475	0.9267	0.9400
F_score	0.8339	0.9125	0.9096	0.9638	0.9319	0.9653
ROC	0.8340	0.9020	0.8930	0.9300	0.9070	0.8360

three well-known ensemble learning techniques (Bagging, Boosting, and Voting), therefore, are applied. These ensemble learning algorithms can improve the performance of classical machine learning algorithms in different domains as can be found in the literature [60], [62]–[68]. To do so, KNN, NB and SVM are combined using such ensemble learning techniques. The obtained outcomes using ensemble learning techniques are illustrated in Figure 3 and Table 4.

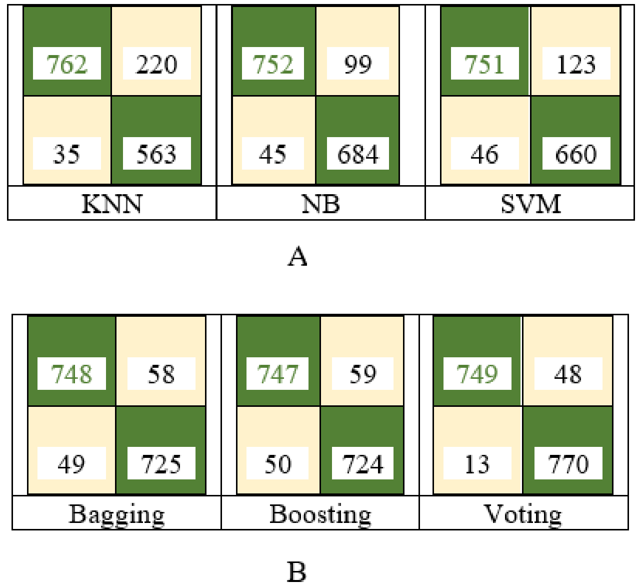


FIGURE 3. Confusion matrices for classical algorithms. A: Confusion matrices for classical algorithms, B: Confusion matrices for combination of ensemble learning techniques and classical algorithms.

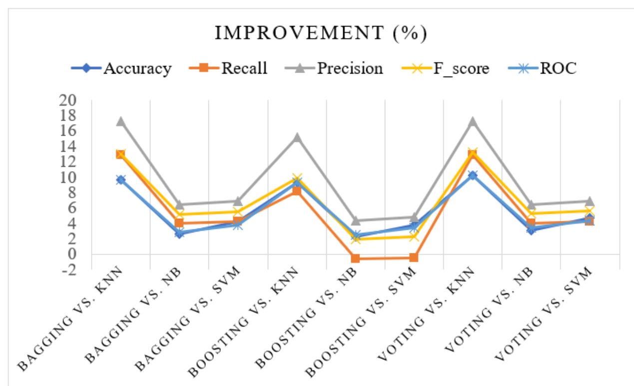


FIGURE 4. Improvement rates after applying ensemble learning techniques compared to simple machine learning algorithms (%).

As it is shown, in terms of accuracy, F_score and ROC, Voting technique has the best performance compared to Bagging and Boosting techniques while Bagging and Voting techniques had the same recall and precision rate values. The amount of optimization (improvement rate (%)) by each ensemble machine learning algorithm compared to each classical algorithm is illustrated by Figure 3. According to Figure 3, it is very clear that the performances of classical algorithms are improved except the performance of Boosting technique versos NB and SVM algorithms in terms of recall with slight decreases per each (-0.63 and -0.50, respectively). However, all other metrics are significantly increased as indicated by Figure 4. Moreover, median AUC of all classical and ensemble learning algorithms are shown in Figure 5.

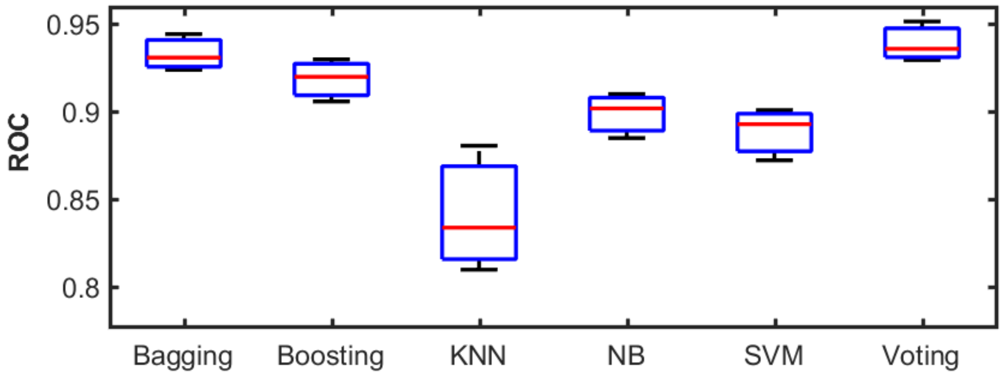


FIGURE 5. Median AUC of the classical algorithms before and after applying ensemble learning techniques.

According to Figure 5, it can be seen that the ensemble learning techniques have better median AUC or (ROC) than classical machine learning algorithms. It should be expressed that voting ensemble learning technique has valuable performance compared to all other methods. An obvious point in Figure 5 is related to KNN which has bigger variation range compared to other methods while other simple methods and ensemble learning techniques have approximately behavior. This shows that the obtained outcomes by 10-fold cross validation in KNN are not very close to each other while other methods have closer results for their 10 folds. Accordingly, we observed that the variation range of optimized methods using ensemble learning techniques are smaller than classical algorithms.

B. SECOND EXPERIMENT

In this step of research, the performance of the proposed methodology will be optimized using four EOAs: The Ant Lion Optimizer (ALO), Whale Optimization Algorithm (WOA), Particle swarm optimization (PSO), and Grey Wolf Optimizer (GWO) algorithms. In this regard, each of these four EOAs will be applied with Bagging, Boosting and Voting separately. It should be noted the combination of ensemble learning techniques with EOAs are called as follows: ALO-Bagging, WOA-Bagging, PSO-Bagging, GWO-Bagging, ALO-Boosting, WOA-Boosting, PSO-Boosting, GWO-Boosting, ALO-Voting, WOA-Voting, PSO-Voting, and GWO-Voting

1) BAGGING

In the first step of this experiment, the performance of Bagging technique is optimized using ALO, WOA, PSO and GWO algorithms. The confusion matrices and measure metrics of this step are indicated by Figure 6 and Table 5. As shown by Bagging in Table 4 and the optimized Bagging using four EOAs in Figure 6, it is obvious that both false positive and false negative rates in all optimized Bagging using EOAs are declined. This means that the error rate of the Bagging algorithm is decreased. On the other hand, it can be observed that the true positive and true negative rates are also increased which means the optimized Bagging can classify both classes batter than previous steps. Moreover, different measures for each EOA is evaluated as presented by Table5. As can be seen, there are different weights by EOAs for each simple algorithm (KNN, SVM, and NB) after using Bagging technique. For example, ALO and PSO give positive weight to KNN whereas WAO and QWO give negative weights. Moreover, an interesting point about SVM is that all EAs give negative weights to this algorithm. Finally, unlike SVM, the positive weights are presented for NB by all EOAs algorithms.

TABLE 5. Measure metrics for hybrid of Bagging and other 4 EOAs: ALO, WOA, PSO and GWO.

Measure	ALO	WAO	PSO	GWO
Max	$9.0958E^{+3}$	$1.9846E^{+3}$	$2.5946E^{+3}$	376.61167
Min	$3.8784E^{-11}$	$2.5657E^{-4}$	4.6151^{-13}	$1.2166E^{-28}$
Mean	18.3871	0.1733	1.5973	8.4319
Median	0.0116	$3.3514E^{-7}$	$1.8114E^{-8}$	$1.0332E^{-5}$
STD	406.7746	0.8171	$1.8114E^{-08}$	36.7045
W_{SVM}	0.5194	-1.6642	0.9689	-1.1416
W_{KNN}	-0.5533	-1.9993	-1.1157	-1.8672
W_{NB}	0.2668	1.3568	0.4513	0.9797

2) BOOSTING

Afterward of Bagging, the performance of Boosting technique is optimized through applying ALO, WOA, PSO and GWO algorithms. The confusion matrices and measure metrics obtained are indicated by and Figure 6 and Table 6

By comparing confusion matrix of Boosting in Figure 3 and the optimized Boosting using four EOAs in Figure 6, the same behavior of the optimized Bagging can be observed in. The final confusion matrix in Figure 6 shows that false positive and false negative rates in all EOAs with Boosting are less than false positive and false negative rates presented in Figure 3. Accordingly, the true positive and true negative rates in this step are much better than previous steps. Hence, it can be argued that the performance of Boosting ensemble learning technique can be enhanced using EOAs. According to Table 6. Unlike KNN in Bagging, here all EOAs provide positive weights for KNN algorithm.

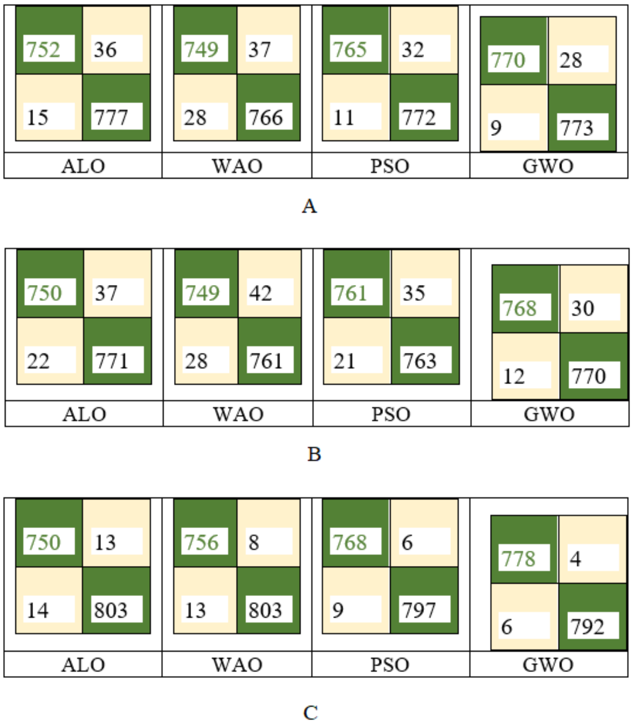


FIGURE 6. Confusion matrices for hybrid of A) Bagging, B) Boosting and C) Voting and other 4 EOAs: ALO, WOA, PSO and GWO.

TABLE 6. Measure metrics for hybrid of Boosting and other 4 EOAs: ALO, WOA, PSO and GWO.

Measure	ALO	WAO	PSO	GWO
Max	$9.0958E^{+5}$	$2.3612E^{+3}$	$92E^{-1}$	$1.0987E^{-1}$
Min	$3.8784E^{-11}$	$5.30091E^{-3}$	$3.5270E^{-12}$	$6.4369E^{-25}$
Mean	1.1997	40.2146	0.64473	402218
Median	$5.0233E^{-5}$	$3.0862E^{-4}$	$2.9834E^{-7}$	$3.5880E^{-13}$
STD	140.4091	296.9496	7.2181	566.4076
W_{SVM}	0.4799	0.1577	1.1320	0.9256
W_{KNN}	0.4727	0.3623	-1.2352	-1.0726
W_{NB}	-1.2358	-0.7341	1.9972	-1.8683

There are different behaviors for SVM and NB in Table 6. SVM gets positive weights by ALO and WAO while PSO and GWO give negative ones. Finally, we can see that NB gains negative weights by all EOAs expect PSO algorithm.

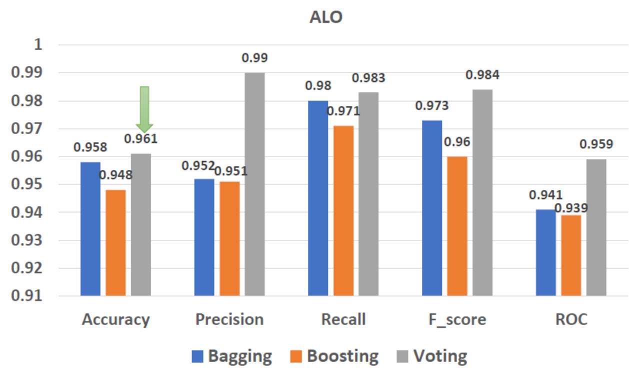
3) VOTING

Finally, all ALO, WOA, PSO and GWO algorithms are used for optimizing the performance of Voting technique. As previously shown, both the confusion matrices and measure metrics are computed in which are presented in Figure 6 and Table 7.

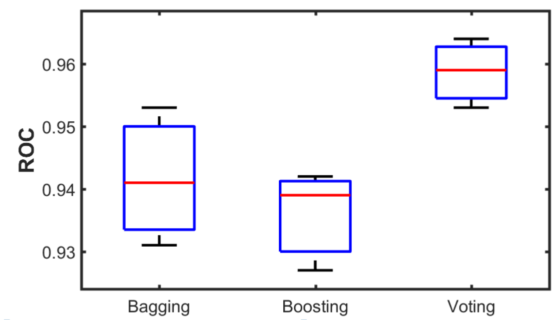
At the end, the performance of Voting ensemble learning technique with four well-known EOAs are investigated. As can be seen in Figure 3 (Voting) and Figure 6, very significant improvements are obvious. A combination of Voting technique with EOAs provide final models with very high true positive and true negative rates compared to single application of Voting technique. This leads to have very low false positive and false negative rates than other methods.

TABLE 7. Measure metrics for hybrid of Voting and other 4 EOAs: ALO, WOA, PSO and GWO.

Measure	ALO	WAO	PSO	GWO
Max	$2.1982E^{+97}$	$2.5953E^{+94}$	$9.7865E^{+95}$	$8.6135E^{+95}$
Min	$4.7592E^{+81}$	1.494e+84E^{+84}	$4.2105E^{81}$	12321
Mean	$4.6699E^{+94}$	$6.9828E^{+92}$	$3.627E^{94}$	$2.7206E^{+95}$
Median	1.2216eE^{+87}	$1.0677E^{+96}$	$1.397E^{+87}$	$3.0176E^{+95}$
STD	$9.8296E^{+96}$	$1.1936E^{+93}$	$5.0926E^{+96}$	$1.9383E^{+95}$
W_{SVM}	-0.1169	0.0859	-0.5004	0.1062
W_{KNN}	0.0137	-0.0074	-0.2802	-0.0113
W_{NB}	0.9322	-2	0.02785	0.5845



(a) Comparison of the ensemble machine learning algorithms and ALO



(b) Median AUC of three ensemble learning algorithms and ALO.

FIGURE 7. Evaluation of performance metrics after applying ALO into the Bagging, Boosting and Voting ensembles.

The measure metrics of Voting, ensemble learning and EOAs are presented by Table 8. According to Table 7, ALO and PSO provide negative weights for KNN whereas WAO and GWO present quite positive ones. Furthermore, it is clear that SVM gains negative weights by all EOAs expect ALO algorithm. Lastly, it can be observed that NB gains positive weights by all EOAs expect WAO algorithm.

Finally, the best score obtained by each EOAs for all Bagging, Boosting and Voting techniques are illustrated in Figure 11. According to Figures 11a, 11b and 11c, GWO showed different behaviors compared to ALO, WOA, and PSO in Bagging, Boosting and Voting techniques. In other words, GWO has very high scores by having less iteration values while it will decrease to zero when the number of iterations in about 350 or higher. Another significant point is that ALO, WOA, and PSO in Voting technique have very similar behavior whereas such behavior cannot be observed in Bagging and Boosting techniques. Finally, we found that the stability of

TABLE 8. Previous research methods applied to driver's behavior analysis (Comparison part).

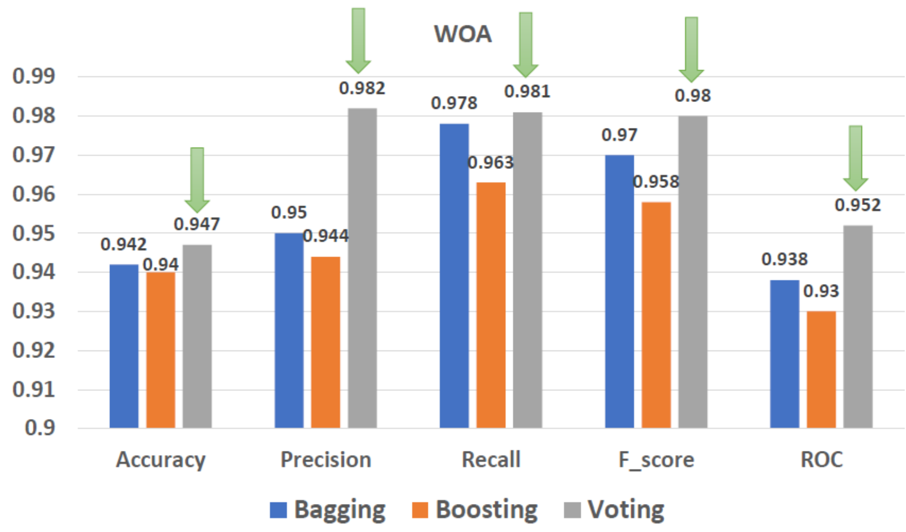
Reference	Type of data	Objective	Method	Accuracy
[43]	Time series data	Driver drowsiness estimation	SBS [12]+SVM	93.00
[74]	Physiological data	Driver drowsiness estimation	ARTE [13] +SVM	93.00 (binary class) 79.00 (for multiclass)
[75]	Time series and vehicle trajectory data from the traffic information	The impact of safety critical events and traffic conditions	SMOTE [14] + TCT [15] + BPNN[16]	88.85 (High-risk and Low-risk)
[80]	Physiological signals	Cross-subject driver status detection	CSDF [17] +MFFS [18] +ARTL [19]	94.44
[76]	RGB videos	Driver drowsiness estimation	Inception V3+FFA [20]	78.54
[78]	EEG signals	Driving fatigue detection	Cross-session validation SVM	83.30
[79]	EEG signals	Driving fatigue detection	ESTCNN [21]	97.37
[77]	EEG signals	Driving fatigue detection	GBDT [22]	94
[72]	Physiological data	Driver drowsiness estimation	RF	81.40
[73]	EEG signals	Driver drowsiness estimation	EOAs+ ELM [23]	87.92
[71]	Physiological data	Driver sleepiness estimation	KSS [24]+ RF	94.10
[70]	EEG signals	Driver drowsiness estimation	LSVM [25]	92
This study	Vehicle information, Physiological signals Biographic features	Driver performance estimation	GWO-Voting of three classifiers (KNN, SVM, NB)	97.50

[12]sequational backward selection
[13]adaption regularization-based transfer learning
[14]synchronization likelihood
[15]traffic conflict technique
[16]back propagation neural network
[17]class sepration and domain fusion
[18]multiple filtering feature selection
[19]adaption regularization-based transfer learning
[20]feature fused architecture
[21]EEG-based spatial temporal convolutional neural network
[22]gradient boosting decision tree
[23]extreme learning machine
[24]karolinska sleepiness scale
[25]linear support vector machine

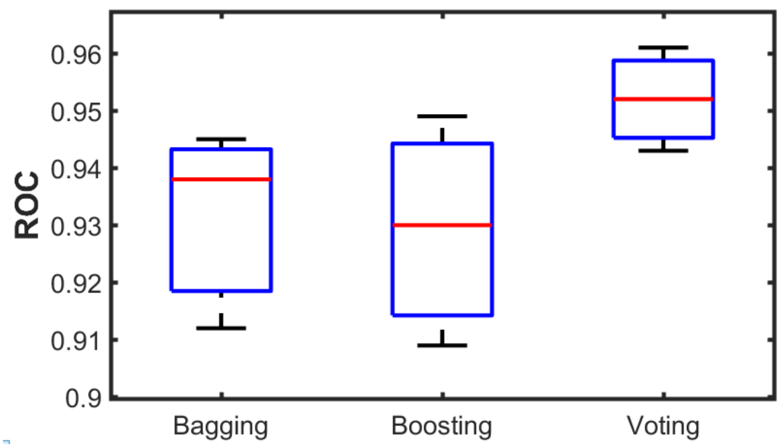
WOA in all ensemble learning techniques (Bagging, Boosting, and Voting) is more than other EA methods.

VII. DISCUSSION

In this phase of study, the obtained results should be compared and discussed in more details. The main point of this study is how combination of ensemble learning techniques and EOAs can be efficient in classification and prediction of driver behavior data. In other words, this study is trying to present a new hybrid model that can be used in real world as an intelligent system. Like previous step, the performance of all ALO-Bagging, WOA-Bagging, PSO-Bagging, GWO-Bagging, ALO-Boosting, WOA-Boosting, PSO-Boosting, GWO-Boosting, ALO-Voting, WOA-Voting, PSO-Voting, and GWO-Voting algorithms will be evaluated using accuracy, precision, recall, F_score and ROC metrics.



(a) Comparison of the ensemble machine learning algorithms and WOA

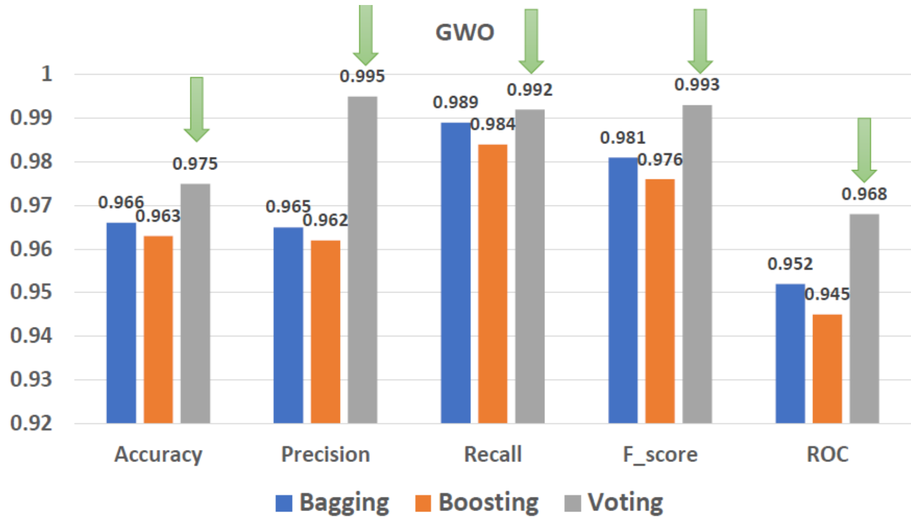


(b) Median AUC of three ensemble learning algorithms and WOA.

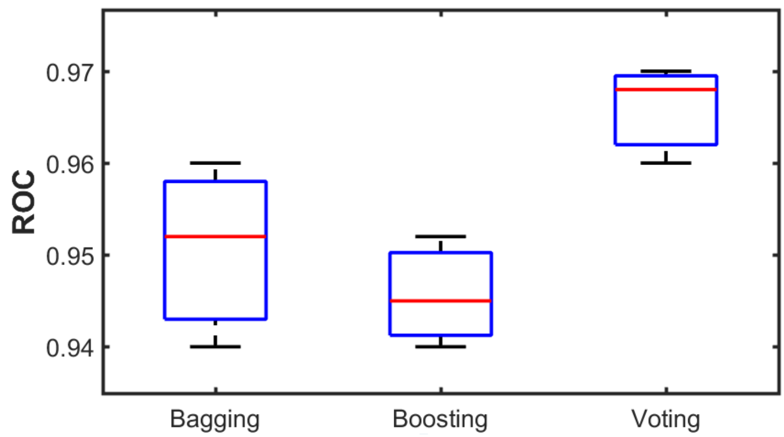
FIGURE 8. **Evaluation of performance metrics after applying WOA into the Bagging, Boosting and Voting ensembles.**

To do so, the performance of each EOA and three ensemble learning techniques will be separately checked here. Furthermore, the median AUC of each step will be also presented. First, the ALO algorithm is applied with Bagging, Boosting and Voting techniques. The performance of ALO with ensemble learning algorithms is shown by Figures 7a and 7b. As shown by Figure 7a we can observe that the best performance is presented by combination of ALO and Voting technique followed by Bagging technique. However, ALO and Boosting had slightly weaker performance compared to other two algorithms. In addition, Figure 7bindicates that the median AUC of Voting in much higher than other two methods. Hence, it can be argued that there is significant difference between Voting technique with Bagging and Boosting techniques. Therefore, it can be argued that the distribution of performance of Bagging for 10 folds are not very close while Boosting and Voting have different outcomes. In other words, we can argue that both Boosting and Voting have almost similar behaviors in 10 folds which is very important for checking final results In the following, the performances of the WOA algorithm with Bagging, Boosting and Voting techniques are illustrated in Figures 8a and 8b.

Like Voting with ALO, Voting with WOA performed outstanding performance compared to Bagging and Boosting. A notable point about Figure 8b is that unlike ALO, the median AUC of WOA with Bagging and Boosting have very close results. As Figure 8b shows, the variation of Bagging is bigger than other methods followed by Boosting technique. A significant point is that Voting technique with WOA has very similar outcomes with ALO (see Figure 7b). This can be considered as a benefit of combing Voting technique and EOAs. Finally, the performance of GWO with three ensemble learning is evaluated in which Figures 9a and 9b represent the outcomes. According to Figure 9a, once again it is obvious that combination of Voting with ensemble learning has better results compared to Bagging and Boosting



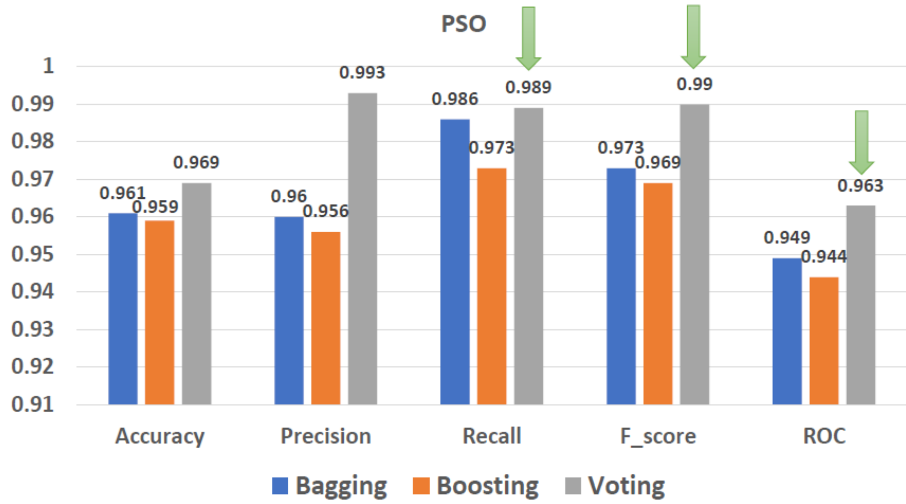
(a) Comparison of the ensemble machine learning algorithms and GWO



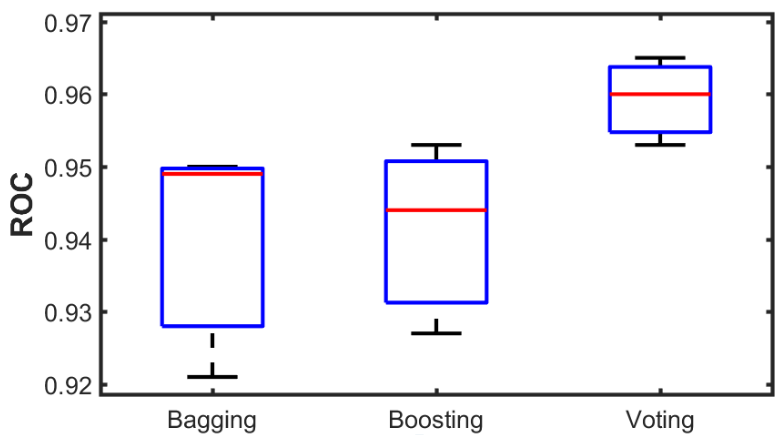
(b) Median AUC of three ensemble learning algorithms and GWO.

FIGURE 9. **Evaluation of performance metrics after applying GWO into the Bagging, Boosting and Voting ensembles.**

techniques. This figure 9b shows that the median AUC of Voting is greater than other two methods. According to this figure, we noticed that the variation range of Bagging with GWO is greater than Boosting and Voting techniques. This is almost similar results with ALO presented in Figure 7b. Thus, it can be argued that the results obtained so far prove that Voting technique with GWO has better results compared to other hybrid methods applied in this study. Accordingly, the performance of the PSO algorithm with three ensemble learning techniques is investigated (see Figure 10a). Once again, we can see that Voting has better outcomes in driver behavior analysis while Bagging and Boosting have very close results together. The behavior of Bagging and Boosting with PSO is almost like WOA. In another words, PSO and WOA have slightly difference in dealing with Bagging and Boosting techniques (see Figures 8a and 10a). As Figure 10b shows, the variation of Bagging is bigger than other methods followed by Boosting technique. A significant point is that Voting technique with WOA has very similar outcomes with ALO (see Figure 7b). This can be considered as a benefit of combing Voting technique and EOAs. According to the outcomes obtained, it can be concluded that Voting technique with all EOAs showed the best performance than other two ensemble learning techniques. Finally, for more clarity, the improvement rates after applying EOA methods with ensemble learning techniques compared to singe ensemble learning techniques is computed as indicated in Figure 12. In this regard, the improvement rate of all optimized methods (ALO-Bagging, WOA-Bagging, PSO-Bagging, GWO-Bagging, ALO-Boosting, WOA-Boosting, PSO-Boosting, GWO-Boosting, ALO-Voting, WOA-Voting, PSO-Voting, and GWO-Voting) will be compared with ensemble learning algorithm applied in the first experiment (see Table 4). According to Figure 12, it is obvious that the performances of most cases are improved using four well-know EOAs. It should be note that the recall value in four cases is slightly decreased, however, the other metrics (accuracy, precision, F_score, and ROC) are significantly enhanced. Generally



(a) Comparison of the ensemble machine learning algorithms and PSO



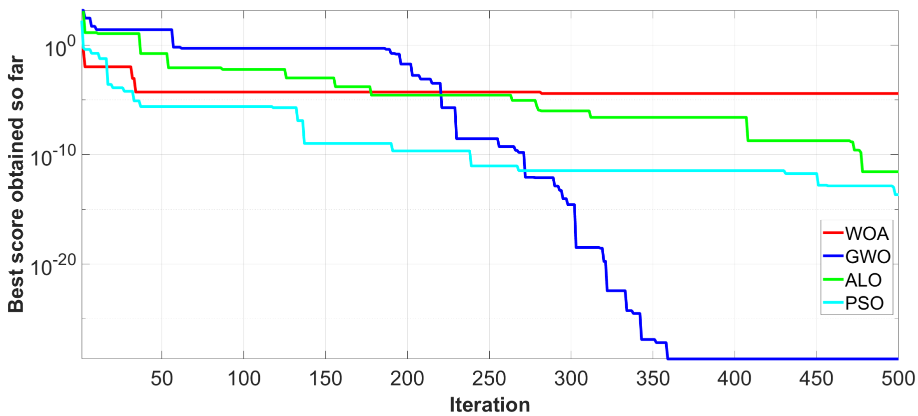
(b) Median AUC of three ensemble learning algorithms and PSO.

FIGURE 10. Evaluation of performance metrics after applying PSO into the Bagging, Boosting and Voting Ensembles.

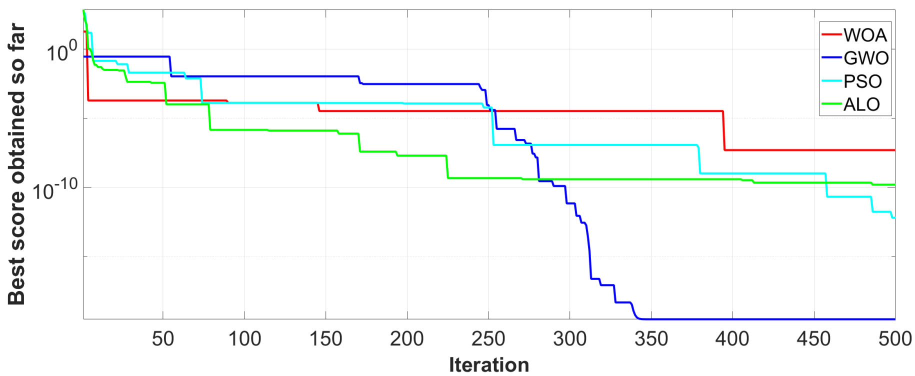
speaking, the proposed integration of ensemble and evolutionary machine learning algorithms for monitoring diver behavior can improve the performance of classical methods, thus, can be implemented as an automated system for detection of performance degradation of drivers. Finally, for more clarity, a comparison of the performance of the proposed methodology is presented by Table 8. Here, the relevant studies in the domain of driver's awareness, drowsiness, and also degradation will be compared.

According to Table 9 it can be seen that our proposed methodology has the best performance and it is weaker than one study only [69]. However, the performance our study is better than other studies presented in Table 8 [70]–[78].

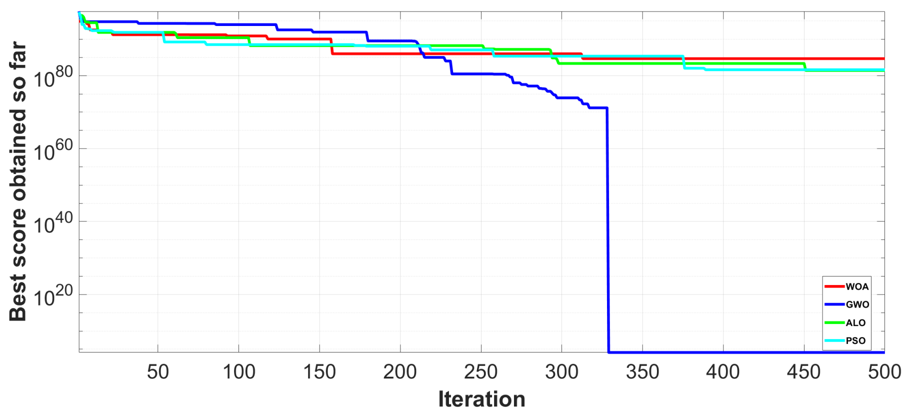
The comprehensive comparison shows that different aspects of driving objectives from different resources have been concluded in the literature. Even though there lots of studies in this domain, however, we report some of the most recent researches. Our comparison indicates that the data used in this study includes three different resources while previous studies considered maximum two resources. Hence this point can be mentioned as a befit of current study. In the following some of those studies with higher accuracy will be discussed. As discussed earlier, [79] investigated Driving fatigue detection using EEG signals and EEG-based spatial–temporal convolutional neural network method. Their proposed method had the accuracy of 97.37 %. Accordingly, [80] applied a machine learning-based system to the Physiological signals in order to detect the cross-subject driver status. In this regard, Class Separation and Domain Fusion approach was used with multiple filtering feature selection and Adaptation Regularization-based Transfer Learning. This hybrid model had 94.44 % accuracy. In future study, we will apply other EOAs to increase the search ability and better convergence. For example, we will apply multi sub-optimizers which are introduced by [24]. This method is applied to enhance the performance of power point tracking (MPPT) under partial shading condition. Other novel methods which are introduced by [32], [81], [82] shows that memetic reinforcement



(a) Bagging Ensemble



(b) Boosting Ensemble



(c) Voting Ensemble

FIGURE 11. The best scores obtained by ALO, WOA, PSO and GWO for (*a*) Bagging (*b*) Boosting (*c*) Voting Ensembles.

learning and grouped grey wolf optimizer can increase the efficiency of MPPT. In [30] Modified Sine Cosine Algorithm (m-SCA) is applied to the common functions to show the effective of this model. There are several Advantages and disadvantages of our proposed methodology. In the following the most important ones will be presented:

Advantages:

- Flexibility in increasing the number of algorithms.
- Significant improvement in performance metrics compared to each single algorithm.
- Considering the effectiveness of each algorithm based on the weight gained from the performance of each algorithm.

Disadvantages:

- Here, we applied only three well-known machine learning algorithms, hence more base algorithms should be investigated.
- In this study, three ensemble learning techniques were applied, whereas more recent ensemble learning should be used.

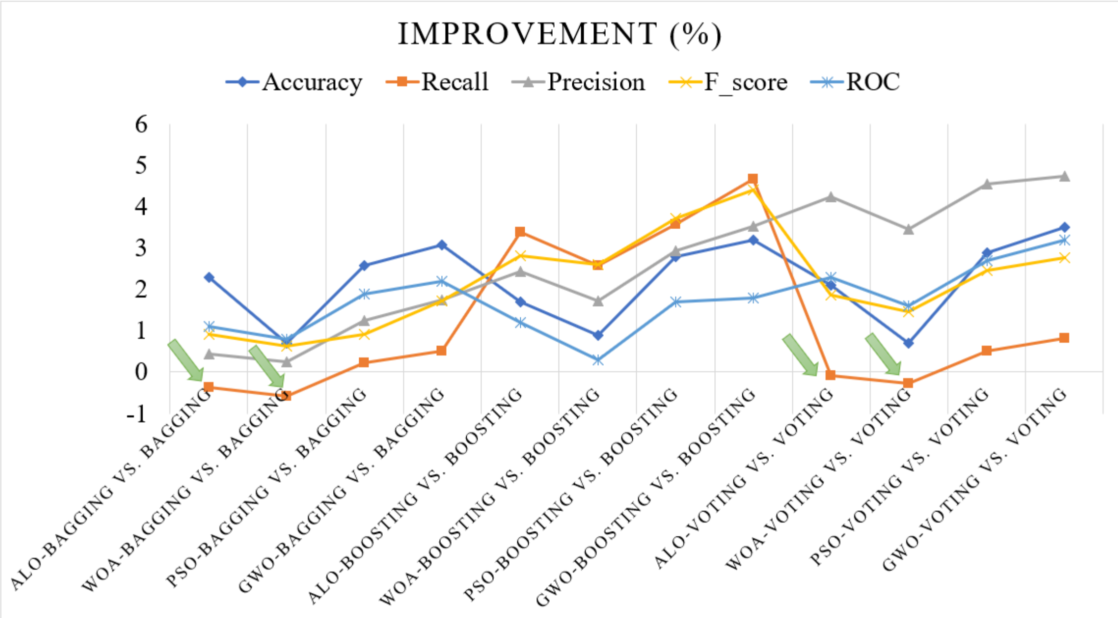


FIGURE 12. Improvement rates after applying EA methods with ensemble learning techniques single ensemble learning techniques (%).

- Moreover, we applied four EOA algorithms, however, since these four algorithms are in their original versions, we would try some new algorithms [24], [30], [32], [81]–[85].

VIII. CONCLUSION

Using a car is one of the most important achievements of the industrial revolution which has had a tremendous effect on the improvement of human life. Despite all these advantages the annual mortality losses caused by the means of transport cannot be ignored. There are different reasons that may affect the performance of drivers which can be grouped into the two main categories: external reasons (e.g., weather, condition and quality of roads, enough warning signs) and internal reasons (e.g., driving fatigue, driver drowsiness, driver degradation). Due to importance of the internal causes, this study, therefore, concentrated on machine learning algorithms to present an automated system for using monitoring diver behavior physiological signals, biographic features and vehicle information collected from 68 subjects. This study, first, applied the three well-known machine learning algorithms: SVM, KNN, and NB. Subsequently, three ensemble learning techniques used to combine these three algorithms for improving their performances. Finally, four EAs (the ALO, WOA, PSO, and GWO algorithms) added to previous system. The proposed methodology could outstandingly improve the performance of used classical algorithms. Among all of them, the GWO-Voting approach had the highest accuracy with 97.50 %. In our future work, we aim to use bigger dataset with more drivers. Moreover, other base classifiers will be investigated. Finally, the performance of another ensemble learning and EAs should be studied.

CONFLICT OF INTEREST

The authors have no conflict of interest to mention.

REFERENCES

[1] N. J. Nilsson, "Introduction to machine learning: An early draft of a proposed textbook," Robot. Lab., Dept. Comput. Sci., Stanford Univ., Stanford, CA, USA, Tech. Rep., 1996.
[2] P. N. Tan, M. Steinbach, and V. Kumar, "Cluster analysis: Additional issues and algorithms," in *Introduction to Data Mining*, 2nd ed. Boston, MA, USA: Addison-Wesley, 2005, pp. 569–650.
[3] P. Baldi, S. Brunak, and F. Bach, *Bioinformatics: The Machine Learning Approach*. Cambridge, MA, USA: MIT Press, 2001.
[4] C. M. Martinez, M. Heucke, F.-Y. Wang, B. Gao, and D. Cao, "Driving style recognition for intelligent vehicle control and advanced driver assistance: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no, 13, pp. 666–676, Mar. 2016.
[5] A. Onan, S. Korukoğlu, and H. Bulut, "A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification," *Expert Syst. Appl.*, vol. 62, pp. 1–16, Nov. 2016.
[6] G. Ridgeway, "The state of boosting," *Comput. Sci. Statist.*, vol. 31, no. 31, pp. 172–181, 1999.
[7] N. C. Oza and K. Tumer, "Classifier ensembles: Select real-world applications," *Inf. Fusion*, vol. 9, pp. 4–20, Jan. 2008.
[8] S. Kumar, J. Ghosh, and M. M. Crawford, "Hierarchical fusion of multiple classifiers for hyperspectral data analysis," *Pattern Anal. Appl.*, vol. 5, no. 2, pp. 210–220, Jun. 2002.
[9] G. Giacinto and F. Roli, "Ensembles of neural networks for soft classification of remote sensing images," in *Proc. Eur. Symp. Intell. Techn.*, 1997, pp. 20–21.
[10] S. Verma, P. Mittal, M. Vatsa, and R. Singh, "At-a-distance person recognition via combining ocular features," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 3131–3135.
[11] H. Erdoğan, A. Erçil, H. K. Ekenel, S. Y. Bilgin, I. Eden, M. Kirişçi, and H. Abut, "Multi-modal person recognition for vehicular applications," in *Proc. Int. Workshop Multiple Classifier Syst.* Berlin, Germany: Springer, 2005, pp. 366–375.

[12] E. Pranckeviciene, R. Baumgartner, and R. Somorjai, "Using domain knowledge in the random subspace method: Application to the classification of biomedical spectra," in *Proc. Int. Workshop Multiple Classifier Syst.* Springer, 2005, pp. 336–345.
[13] S. Mirjalili, G.-G. Wang, and L. D. S. Coelho, "Binary optimization using hybrid particle swarm optimization and gravitational search algorithm," *Neural Comput. Appl.*, vol. 25, no. 6, pp. 1423–1435, 2014.
[14] S. Mirjalili, "The ant lion optimizer," *Adv. Eng. Softw.*, vol. 83, pp. 80–98, May 2015.
[15] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Adv. Eng. Softw.*, vol. 69, pp. 46–61, Mar. 2014.
[16] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Adv. Eng. Softw.*, vol. 95, pp. 51–67, May 2016.
[17] E. Alba and B. Dorronsoro, "The exploration/exploitation tradeoff in dynamic cellular genetic algorithms," *IEEE Trans. Evol. Comput.*, vol. 9, no. 2, pp. 126–142, Apr. 2005.
[18] M. Abdar, M. Zomorodi-Moghadam, and X. Zhou, "An ensemble-based decision tree approach for educational data mining," in *Proc. 5th Int. Conf. Behav., Econ., Socio-Cultural Comput. (BESC)*, Nov. 2019, pp. 126–129.
[19] A. Madadi and M. M. Motlagh, "Optimal control of DC motor using grey wolf optimizer algorithm," *Tech. J. Eng. Appl. Sci.*, vol. 4, no. 4, pp. 373–379, 2014.
[20] X. Song, L. Tang, S. Zhao, X. Zhang, L. Li, J. Huang, and W. Cai, "Grey Wolf Optimizer for parameter estimation in surface waves," *Soil Dyn. Earthquake Eng.*, vol. 75, pp. 147–157, Aug. 2015.
[21] M. R. Shakarami and I. F. Davoudkhani, "Wide-area power system stabilizer design based on grey wolf optimization algorithm considering the time delay," *Electr. Power Syst. Res.*, vol. 133, pp. 149–159, Apr. 2016.
[22] D. Guha, P. K. Roy, and S. Banerjee, "Load frequency control of interconnected power system using grey wolf optimization," *Swarm Evol. Comput.*, vol. 27, pp. 97–115, Apr. 2016.
[23] R. Sindhu, R. Ngadiran, Y. M. Yacob, N. A. H. Zahri, M. Hariharan, and K. Polat, "A hybrid SCA inspired BBO for feature selection problems," *Math. Problems Eng.*, vol. 2019, 2019, Art. no. 9517568.
[24] B. Yang, T. Yu, X. Zhang, H. Li, H. Shu, Y. Sang, and L. Jiang, "Dynamic leader based collective intelligence for maximum power point tracking of PV systems affected by partial shading condition," *Energy Convers. Manage.*, vol. 179, pp. 286–303, Jan. 2019.
[25] B. Yang, L. Zhong, X. Zhang, H. Shu, T. Yu, H. Li, L. Jiang, and L. Sun, "Novel bio-inspired memetic salp swarm algorithm and application to MPPT for PV systems considering partial shading condition," *J. Cleaner Prod.*, vol. 215, pp. 1203–1222, Apr. 2019.
[26] Z. Arabasadi, R. Alizadehsani, M. Roshanzamir, H. Moosaei, and A. A. Yarifard, "Computer aided decision making for heart disease detection using hybrid neural network-genetic algorithm," *Comput. Methods Programs Biomed.*, vol. 141, pp. 19–26, Apr. 2017.
[27] I. Zohdy and H. Rakha, "Optimizing driverless vehicles at intersections," in *Proc. 10th ITS World Congr.*, Vienna, Austria, 2012, pp. 1–12.
[28] W. Long, T. Wu, S. Cai, X. Liang, J. Jiao, and M. Xu, "A novel grey wolf optimizer algorithm with refraction learning," *IEEE Access*, vol. 7, pp. 57805–57819, 2019.
[29] P. Hu, S. Chen, H. Huang, G. Zhang, and L. Liu, "Improved alpha-guided Grey wolf optimizer," *IEEE Access*, vol. 7, pp. 5421–5437, Jan. 2019.
[30] S. Gupta and K. Deep, "A hybrid self-adaptive sine cosine algorithm with opposition based learning," *Expert Syst. Appl.*, vol. 119, pp. 210–230, Apr. 2019.
[31] S. Gupta and K. Deep, "A novel random walk grey wolf optimizer," *Swarm Evol. Comput.*, vol. 44, pp. 101–112, Feb. 2019.
[32] B. Yang, X. Zhang, T. Yu, H. Shu, and Z. Fang, "Grouped grey wolf optimizer for maximum power point tracking of doubly-fed induction generator based wind turbine," *Energy Convers. Manage.*, vol. 133, pp. 427–443, Feb. 2017.
[33] W. G. Najm, J. D. Smith, and M. Yanagisawa, "Pre-crash scenario typology for crash avoidance research," Nat. Tech. Inf. Service, Springfield, VA, USA, Tech. Rep., 2007.
[34] *Traffic Safety Facts 2015*, National Highway Traffic Safety Administration, Washington, DC, USA, 2016.
[35] Y. Tanaka, T. Bando, M. Egawa, H. Okuda, H. Terai, T. Hirayama, C. Miyajima, D. Deguchi, K. Kaji, and K. Takeda, "Toward the development of a driving support system for repressing overtrust and overreliance," in *Proc. 20th ITS World Congr.*, 2013, pp. 1–10.
[36] D. Shrestha, D. J. Lovell, and Y. Tripodis, "Hardware and software for collecting microscopic trajectory data on naturalistic driving behavior," *J. Intell. Transp. Syst.*, vol. 21, no. 3, pp. 202–213, 2017.
[37] W. Enkelmann, "Video-based driver assistance–from basic functions to applications," *Int. J. Comput. Vis.*, vol. 45, no. 3, pp. 201–221, 2001.
[38] R. Risack, N. Mohler, and W. Enkelmann, "A video-based lane keeping assistant," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Oct. 2000, pp. 356–361.
[39] B. Fardi, U. Scheunert, H. Cramer, and G. Wanielik, "A new approach for lane departure identification," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2003, pp. 100–105.
[40] D. L. Strayer, J. M. Cooper, J. Turrill, J. Coleman, N. Medeiros-Ward, and F. Biondi, "Measuring cognitive distraction in the automobile," Tech. Rep., 2013.
[41] S. Y. Cheng, S. Park, and M. M. Trivedi, "Multi-spectral and multi-perspective video arrays for driver body tracking and activity analysis," *Comput. Vis. Image Understand.*, vol. 106, nos. 2–3, pp. 245–257, 2007.
[42] Y. Lin, H. Leng, G. Yang, and H. Cai, "An intelligent noninvasive sensor for driver pulse wave measurement," *IEEE Sensors J.*, vol. 7, no. 5, pp. 790–799, May 2007.
[43] B. Chakraborty and K. Nakano, "Automatic detection of driver's awareness with cognitive task from driving behavior," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2016, pp. 3630–3633.
[44] J. Yu, Z. Chen, Y. Zhu, Y. J. Chen, L. Kong, and M. Li, "Fine-grained abnormal driving behaviors detection and identification with smartphones," *IEEE Trans. Mobile Comput.*, vol. 16, no. 8, pp. 2198–2212, Oct. 2017.
[45] Y. Wang, X. Liu, Y. Zhang, Z. Zhu, D. Liu, and J. Sun, "Driving fatigue detection based on EEG signal," in *Proc. 5th Int. Conf. Instrum. Meas., Comput., Commun. Control (IMCCC)*, Sep. 2015, pp. 715–718.
[46] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 156–166, Jun. 2005.
[47] A. D. McDonald, J. D. Lee, C. Schwarz, and T. L. Brown, "Steering in a random forest: Ensemble learning for detecting drowsiness-related lane departures," *Human Factors*, vol. 56, no. 5, pp. 986–998, Aug. 2014.
[48] C. D. Katsis, N. Katertsidis, G. Ganiatsas, and D. I. Fotiadis, "Toward emotion recognition in car-racing drivers: A biosignal processing approach," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 38, no. 3, pp. 502–512, May 2008.
[49] H. Eren, S. Makinist, E. Akin, and A. Yilmaz, "Estimating driving behavior by a smartphone," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2012, pp. 234–239.
[50] B. A. Shiferaw, L. A. Downey, J. Westlake, B. Stevens, S. M. Rajaratnam, D. J. Berlowitz, P. Swann, and M. E. Howard, "Stationary gaze entropy predicts lane departure events in sleep-deprived drivers," *Sci. Rep.*, vol. 8, no. 1, 2018, Art. no. 2220.
[51] S. Taamneh, P. Tsiamyrtzis, M. Dcosta, P. Buddharaju, A. Khatri, M. Manser, T. Ferris, R. Wunderlich, and I. Pavlidis, "A multimodal dataset for various forms of distracted driving," *Sci. Data*, vol. 4, Aug. 2017, Art. no. 170110.
[52] J. Goodenough, B. McGuire, and E. Jakob, *Perspectives on Animal Behavior*. Hoboken, NJ, USA: Wiley, 2009.
[53] S. Mirjalili, P. Jangir, and S. Saremi, "Multi-objective ant lion optimizer: A multi-objective optimization algorithm for solving engineering problems," *Appl. Intell.*, vol. 46, no. 1, pp. 79–95, Jan. 2017.
[54] S. Mirjalili, S. Saremi, S. M. Mirjalili, and L. D. S. Coelho, "Multi-objective grey wolf optimizer: A novel algorithm for multi-criterion optimization," *Expert Syst. Appl.*, vol. 47, pp. 106–119, Apr. 2016.
[55] W. A. Watkins and W. E. Schevill, "Aerial observation of feeding behavior in four baleen whales: *Eubalaena glacialis, balaenoptera borealis, megaptera novaeangliae, and balaenoptera physalus*," *J. Mammal.*, vol. 60, no. 1, pp. 155–163, 1979.
[56] S. B. Taieb, G. Bontempi, A. F. Atiya, and A. Sorjamaa, "A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition," *Expert Syst. Appl.*, vol. 39, no. 8, pp. 7067–7083, Jun. 2012.
[57] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems?" *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3133–3181, 2014.
[58] D. Coomans and D. L. Massart, "Alternative k-nearest neighbour rules in supervised pattern recognition: Part 1. K-nearest neighbour classification by using alternative voting rules," *Anal. Chim. Acta*, vol. 136, pp. 15–27, Jan. 1982.
[59] J. Han and M. Kamber, *Introduction: Data Mining. Concepts and Techniques*. San Francisco, CA, USA: Elsevier, 2006, ch. 1.

[60] M. Abdar, N. Y. Yen, and J. C. Hung, "Educational data mining based on multi-objective weighted voting ensemble classifier," in *Proc. Int. Conf. Comput. Sci. Comput. Intell. (CSCI)*, Dec. 2017, pp. 357–362.

[61] J. J. Rodriguez, L. I. Kuncheva, and C. J. Alonso, "Rotation forest: A new classifier ensemble method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1619–1630, Oct. 2006.

[62] M. Hassoon, M. S. Kouhi, M. Zomorodi-Moghadam, and M. Abdar, "Rule optimization of boosted C5.0 classification using genetic algorithm for liver disease prediction," in *Proc. Int. Conf. Comput. Appl. (ICCA)*, Sep. 2017, pp. 299–305.

[63] Q. Fan and X. Zhong, "A triangle voting algorithm based on double feature constraints for star sensors," *Adv. Space Res.*, vol. 61, no. 4, pp. 1132–1142, 2018.

[64] M. Abdar, N. Y. Yen, and J. C.-S. Hung, "Improving the diagnosis of liver disease using multilayer perceptron neural network and boosted decision trees," *J. Med. Biol. Eng.*, vol. 38, no. 6, pp. 953–965, 2017.

[65] J. Sun, J. Lang, H. Fujita, and H. Li, "Imbalanced enterprise credit evaluation with DTE-SBD: Decision tree ensemble based on SMOTE and bagging with differentiated sampling rates," *Inf. Sci.*, vol. 425, pp. 76–91, Jan. 2018.

[66] M. Abdar, M. Zomorodi-Moghadam, X. Zhou, R. Gururajan, X. Tao, P. D. Barua, and R. Gururajan, "A new nested ensemble technique for automated diagnosis of breast cancer," *Pattern Recognit. Lett.*, to be published.

[67] A. Andiojaya and H. Demirhan, "A bagging algorithm for the imputation of missing values in time series," *Expert Syst. Appl.*, vol. 129, pp. 10–26, Sep. 2019.

[68] R. Lin, Z. Pei, Z. Ye, B. Wu, and G. Yang, "A voted based random forests algorithm for smart grid distribution network faults prediction," *Enterprise Inf. Syst.*, to be published.

[69] J. Chen, H. Wang, and C. Hua, "Assessment of driver drowsiness using electroencephalogram signals based on multiple functional brain networks," *Int. J. Psychophysiol.*, vol. 133, pp. 120–130, Nov. 2018.

[70] A. Mehreen, S. M. Anwar, M. Haseeb, M. Majid, and M. O. Ullah, "A hybrid scheme for drowsiness detection using wearable sensors," *IEEE Sensors J.*, vol. 19, no. 13, pp. 5119–5126, Mar. 2019.

[71] H. Mårtensson, O. Keelan, and C. Ahlström, "Driver sleepiness classification based on physiological data and driving performance from real road driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 2, pp. 421–430, Apr. 2018.

[72] J. Gwak, M. Shino, and A. Hirao, "Early detection of driver drowsiness utilizing machine learning based on physiological signals, behavioral measures, and driving performance," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 1794–1800.

[73] S. Taran and V. Bajaj, "Drowsiness detection using adaptive Hermite decomposition and extreme learning machine for electroencephalogram signals," *IEEE Sensors J.*, vol. 18, no. 21, pp. 8855–8862, Nov. 2018.

[74] S. Barua, "Multivariate data analytics to identify driver's sleepiness, cognitive load, and stress," Ph.D. dissertation, Mälardalen Univ., Västerås, Sweden, 2019.

[75] J. Wang, Y. Kong, and T. Fu, "Expressway crash risk prediction using back propagation neural network: A brief investigation on safety resilience," *Accident Anal. Prevention*, vol. 124, pp. 180–192, Mar. 2019.

[76] V. Vijayan and E. Sherly, "Real time detection system of driver drowsiness based on representation learning using deep neural networks," *J. Intell. Fuzzy Syst.*, to be published.

[77] J. Hu and J. Min, "Automated detection of driver fatigue based on EEG signals using gradient boosting decision tree model," *Cogn. Neurodyn.*, vol. 12, no. 4, pp. 431–440, 2018.

[78] C.-S. Wei, Y.-T. Wang, C.-T. Lin, and T.-P. Jung, "Toward drowsiness detection using non-hair-bearing eeg-based brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 2, pp. 400–406, Feb. 2018.

[79] Z. Gao, X. Wang, Y. Yang, C. Mu, Q. Cai, W. Dang, and S. Zuo, "EEG-based spatio-temporal convolutional neural network for driver fatigue evaluation," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published.

[80] L.-L. Chen, A. Zhang, and X.-G. Lou, "Cross-subject driver status detection from physiological signals based on hybrid feature selection and transfer learning," *Expert Syst. Appl.*, to be published.

[81] X. Zhang, S. Li, T. He, B. Yang, T. Yu, H. Li, L. Jiang, and L. Sun, "Memetic reinforcement learning based maximum power point tracking design for PV systems under partial shading condition," *Energy*, vol. 174, pp. 1079–1090, May 2019.

[82] B. Yang, T. Yu, H. Shu, D. Zhu, N. An, Y. Sang, and L. Jiang, "Energy reshaping based passive fractional-order PID control design and implementation of a grid-connected PV inverter for MPPT using grouped grey wolf optimizer," *Sol. Energy*, vol. 170, pp. 31–46, Aug. 2018.

[83] A. S. Shamsaldin, T. A. Rashid, R. A. A.-R. Agha, N. K. Al-Salihi, and M. Mohammadi, "Donkey and smuggler optimization algorithm: A collaborative working approach to path finding," *J. Comput. Des. Eng.*, to be published.

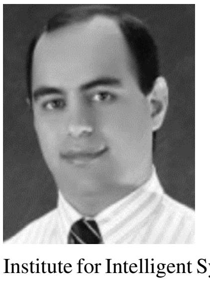
[84] J. M. Abdullah and T. A. Rashid, "Fitness dependent optimizer: Inspired by the bee swarming reproductive process," *IEEE Access*, vol. 7, pp. 43473–43486, 2019.

[85] Q. Tu, X. Chen, and X. Liu, "Multi-strategy ensemble grey wolf optimizer and its application to feature selection," *Appl. Soft Comput.*, vol. 76, pp. 16–30, Mar. 2019.

AFSANEH KOOHESTANI is currently pursuing the Ph.D. degree with Deakin University, Waurn Ponds Campus, Waurn Ponds, VIC, Australia, under the supervision of Dr. A. Khosravi. Her current research interests include data mining and machine learning.

MOLOUD ABDAR received the bachelor's degree in computer engineering from Damghan University, Iran, in 2015, and the master's degree in computer science and engineering from the University of Aizu, Aizu, Japan, in 2018. He is currently pursuing the Ph.D. degree with the University of Montreal, Montreal, QC, Canada. He has several papers about data mining, machine learning, and user modeling in some refereed international journals and conferences. His research interests include data mining, machine learning, ensemble learning, evolutionary algorithms, and user modeling.

He is a member of the International Association of Engineers. He was a recipient of the Fonds de Recherche du Quebec–Nature et Technologies Award, in 2019. He is also very active in five international conferences, such as IEEE AINA 2018 and IEEE AINA 2019, and some referred international journals, such as the IEEE Access, the *Future Generation Computer Systems* (Outstanding Reviewer in October 2017), the *Neurocomputing* (Outstanding Reviewer in January 2017), the *Physica A: Statistical Mechanics and Its Applications*, the *Computers in Human Behavior*, the *International Journal of Applied Mathematics and Computer Science*, the *Expert Systems*, the *Journal of Internet Technology*, the *Health Information Management Journal*, the *Neural Computing and Applications*, the *Journal of Medical Systems*, the *Computer Methods and Programs in Biomedicine*, and the *Web Intelligence* as a Reviewer.



ABBAS KHOSRAVI (M'10) received the B.Sc. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2002, and the M.Sc. degree (Hons.) in electrical engineering from the Amirkabir University of Technology, Tehran, in 2005. He joined as a Research Academic with the eXiT Group, University of Girona, Girona, Spain, in 2006, and researching in the area of artificial intelligence (AI) applications. He is currently a Senior Research Fellow with the Institute for Intelligent Systems Research and Innovation, Deakin University, Waurn Ponds Campus, Waurn Ponds, VIC, Australia. His current research interests include the development and application of AI techniques for (meta) modeling, analysis, control, and optimization of operations within complex systems.

SAEID NAHAVANDI (M'92–SM'07) received the Ph.D. degree from Durham University, Durham, U.K., in 1991. He is currently an Alfred Deakin Professor, the Pro Vice-Chancellor (Defence Technologies), the Chair of Engineering, and the Director of the Institute for Intelligent Systems Research and Innovation, Deakin University, Waurn Ponds Campus, Waurn Ponds, VIC, Australia. He has published over 600 papers in various international journals and conferences. His current research interests include modeling of complex systems, robotics, and haptics. He is a fellow of the Engineers Australia and the Institution of Engineering and Technology. He is the Co-Editor-in-Chief of the IEEE SYSTEMS JOURNAL, an Associate Editor of the IEEE/ASME TRANSACTIONS ON MECHATRONICS and the IEEE TRANSACTIONS ON SYSTEMS, MAN AND CYBERNETICS: SYSTEMS, and an Editorial Board Member of the IEEE ACCESS.

MAHEREH KOOHESTANI received the M.Sc. degree in computer engineering-control from the Department of Electrical Engineering, Islamic Azad University of Mashhad, Mashhad, Iran, in 2018.