International Journal of Engineering and Technical Research (IJETR)
ISSN: 2321-0869 (O) 2454-4698 (P), Volume-5, Issue-2, June 2016

A Hybrid Ensemble Classification Algorithm Using Grey Wolf Optimizer for Terrorism Prediction

Ghada M. A. Soliman, Motaz M. H. Khorshid, Tarek H. M. Abou-El-Enien

Abstract **Grey Wolf Optimizer (GWO) is a recently developed meta-heuristic search algorithm inspired by grey wolves (Canis lupus), which simulate the social stratum and hunting mechanism of grey wolves in nature and based on three main steps of hunting: searching for prey, encircling prey and attacking prey. This paper proposes a hybrid optimized ensemble classification algorithm for terrorism prediction. The proposed algorithm implements grey wolf optimizer (GWO) and wrapper feature selection approach in order to select optimal feature subset for classification process based on random forests (RFs) ensemble classifier to improve and enhance the classification accuracy while minimizing the number of selected features. The performance of the hybrid GWO-RFs algorithm is tested by two different experiments during 20 iterations and the results are benchmarked for evaluation with particle swarm optimization (PSO) and genetic algorithm (GA) with multi-parent recombination, as well as the results of RF classifier are compared with another well known classifier as K-nearest neighbor (KNN). A set of assessment indicators are used to evaluate and compare between the obtained results which prove the capability of the proposed hybrid GWO-RFs algorithm to search the feature space for the optimal feature combination as well as enhancing the classification accuracy compared to other well-known conventional, heuristics and meta-heuristics search algorithms. Experimental results demonstrate competitive performance of the proposed Hybrid GWO-RF ensemble prediction classification algorithm, especially with high dimension datasets**

Index **Meta-Heuristic, Swarm Intelligence, Grey Wolf Optimization, Feature Selection.**

I. INTRODUCTION

Nature-inspired algorithms are becoming popular over the last decades and among researchers due to their simplicity and flexibility. The nature-inspired meta-heuristic algorithms are analyzed in terms of their key features like their diversity and adaptation, exploration and exploitation, and attractions and diffusion mechanisms. The success and challenges concerning these algorithms are based on their parameter tuning and parameter control. Meta-heuristic extended to cover many different areas of study. Surprisingly, some of them such as Genetic Algorithm (GA) [1], Ant Colony Optimization (ACO) [2], and Particle Swarm Optimization (PSO) [3], Differential Evolution (DE) [4], Evolutionary Strategy (ES) [5], and Evolutionary Programming (EP) [6] are fairly well-known among not only computer scientists but also scientists from different fields and have many applications in different branches of science and industry as well. As the complexity of the problems increases over the last few decades, the need for new optimization techniques becomes evident more than before and according to No-Free-Lunch (NFL) theorem [7], there is no algorithm for solving all optimization problems. In other words, the average performance of optimizers is equal when considering all optimization problems [8]. Therefore there are still problems that can be solved by new optimizers better than the current optimizers. Grey Wolf Optimizer (GWO) is a new swarm intelligent (SI) population-based meta heuristic which employed to solve optimization problems of different varies [9]. GWO is a mathematical model and the computer simulation which mimics the leadership hierarchy and hunting mechanism of grey wolves in nature.

Nowadays, Machine Learning (ML) techniques play a very significant role in solving different classification, analysis, and forecasting problems in several areas [10]. One of the most important tasks is classification which is the process of classifying data into predefined categories (classes) based on their content [11]. Supervised Machine learning classification is one of the tasks most frequently carried out by so called Intelligent Systems. Thus, a large number of techniques have been developed based on Artificial Intelligence (Logic-based techniques, Perceptron-based techniques) and Statistics (Bayesian Networks, Instance-based techniques).

The concept of combining classifiers (ensemble methods) is proposed as a new direction for the improvement of the performance of individual machine learning algorithms, and have attracted a great attention of the scientific community over the last years. Hybrid and ensemble methods in machine learning are learning algorithms that construct a set of many individual classifiers (called base learners) and combine them to classify new data points by taking a weighted or unweighted vote of their predictions [12]. Multiple, ensemble learning models have been theoretically and empirically shown to provide significantly better performance than single weak learners, especially while dealing with high dimensional, complex regression and classification problems [13]. The Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. Random forests have been shown to give excellent performance on a number of practical problems. They work fast, generally exhibit a substantial performance improvement over single tree classifiers such as CART, and yield generalization error rates that compare favorably to the best statistical and machine learning methods. In fact, random forests are among the most accurate general-purpose classifiers available [11].

In this research study, the proposed hybrid GWO-RF model implements grey wolf optimizer (GWO) and wrapper feature

Ghada M. A. Soliman, DS & OR Department , Faculty of Computers and Information, Cairo University, Egypt.

Motaz M. H. Korshid, DS & OR Department , Faculty of Computers and Information, Cairo University, Egypt.

Tarek H. M. Abou-El-Enien, DS & OR Department, Faculty of Computers and Information, Cairo University, Egypt.

selection approach in order to select optimal feature subset for classification process based on random forests (RFs) ensemble classifier to improve and enhance the classification accuracy while minimizing the number of selected features. The obtained experimental results indicate significant enhancements in terms of classification accuracy compared with other known meta-heuristics like GA and PSO, as well as the results compared with a hybrid GOW-KNN (K-Nearest Neighbour) classification algorithm to show the superiority of using ensemble classifier among other classification algorithms.

The remainder of this paper is organized as follows. Section II provides background information. Section III describes ensemble learning methods and different algorithms with a focus on Random Forests (RF) classifier. Section IV presents the Feature Selection concept, different techniques and approaches used in this area with a detailed illustration of Grey Wolf Optimizer (GWO) as one of most recent meta-heuristics algorithm proved high performance in that area. Section V explains in details the proposed hybrid prediction classification system. Section VI presents the experimental results and analysis of the proposed system. Section VII provides Conclusions and future work.

II. LITERATURE REVIEW

Throughout the years, multiple techniques for feature selection have been proposed. Some famous FS approaches are based on the Genetic Algorithm [14], Simulated Annealing (SA), Particle Swarm Optimization [15] and Ant Colony Optimization (ACO) [16]-[17]. Among many FS techniques, GA-based methods and ACO- based methods have been attracted a lot of attention, these methods attempts to achieve better solution by using knowledge from previous iterations [18]. PSO algorithm has been applied to random forest classifiers in order to weight the classes' scores as explained in [19].

Greedy search based on sequential backward selection (SBS) [20] and sequential forward selection (SFS) [21] are two model wrapper techniques. SBS (SFS) starts with all attributes (no attributes), then candidate attributes are consecutively removed to (added from) the subset till the further removal (addition) does not rise the classification accuracy. But, these two techniques suffer from the issue of so-called nesting effect, that means once an attribute is eliminated (chosen) it could not be chosen (eliminated) later. This issue could be resolved by merging both SFS and SBS into one technique. Thus, Stearns in [22] proposes a plus-*l*-take away-*k* technique, which performs *l* times forward selection followed by *k* times backward elimination. However, it is hard to detect the best magnitudes of (*l*, *k*).
FOCUS in [23] is a filter attribute reduction technique, which exhaustively examines all potential attribute subsets and then chooses the minimal attribute subset. But, the FOCUS technique was not computationally efficient due to the exhaustive search.

III. ENSEMBLE LEARNING

Ensemble methods popular in machine learning, are learning algorithms that construct a set of many individual classifiers (called base learners) and combine them to classify new data points by taking a weighted or unweighted vote of their predictions. It is now well-known that ensembles are often much more accurate than the individual classifiers that make them up. The success of ensemble algorithms on many benchmark data sets has raised considerable interest in understanding why such methods succeed and identifying circumstances in which they can be expected to produce good results. These methods differ in the way the base learner is fit and combined. For example, bagging by Breiman [24] proceeds by generating bootstrap samples from the original data set, constructing a classifier from each bootstrap sample, and voting to combine. In boosting by Freund and Schapire [25] and arcing algorithms by Breiman [26], the successive classifiers are constructed by giving increased weight to those points that have been frequently misclassified, and the classifiers are combined using weighted voting. On the other hand, random split selection by Dietterich [27], Breiman [18] provides a general framework for tree ensembles called “random forests”. Each tree depends on the values of a random vector sampled independently and with the same distribution for all trees. Thus, a random forest is a classifier that consists of many decision trees and outputs the class that is the mode of the classes output by individual trees. Algorithms for inducing a random forest were first developed by Breiman and Cutler.

Ensemble methods are learning algorithms that construct a set of classifier and then classify new data points by taking (weighted) vote by their predictions [1]. An ensemble of classifiers is a set of classifiers whose individual decisions are combined in some way typically by weighted or unweighted voting to classify new examples. One of the most active areas of research in supervised learning has been to study methods for constructing good ensembles of classifiers. The main discovery is that ensembles are often much more accurate than the individual classifiers that make them up.

A necessary and sufficient condition for an ensemble of classifiers to be more accurate than any of its individual members is to be accurate and diverse [2]. An accurate classifier is one that has an error rate of better than random guessing on new x values.

A. Random Forests

The Random Forests (RF) is one of the best known classification and regression techniques, which has the ability to classify large dataset with excellent accuracy. Random forest classifier is an ensemble classifier that consists of several decision trees [28]. The output of this classifier is the class number that most frequently occurs individually in the output of decision trees classifiers. The main idea of decision trees is to predicate a target based on a group of input data. Decision trees also named classification trees, where the tree leaves represent the class labels and the branches represent the conjunction of feature vectors that lead to class labels.

Random forests have been shown to give excellent performance on a number of practical problems. They work fast, generally exhibit a substantial performance improvement over single tree classifiers such as CART, and yield generalization error rates that compare favorably to the best statistical and machine learning methods. In fact, random forests are among the most accurate general-purpose classifiers available [27]. Different random forests differ in how randomness is introduced in the tree building process, ranging from extreme random splitting strategies [30]-[31] to more involved data-dependent strategies [32]-[27]. As a matter of fact, the statistical mechanism of random forests is not yet fully understood and is still under active investigation. Unlike single trees, where consistency is proved letting the number of observations in each terminal node become large [33], random forests are generally built to have a small number of cases in each terminal node. Although the mechanism of random forest algorithms appears simple, it is difficult to analyze and remains largely unknown. Some attempts to investigate the driving force behind consistency of random forests are by [33]-[35] who establish a connection between random forests and adaptive nearest neighbor methods. Meinshausen in [36] proved consistency of certain random forests in the context of so-called quantile regression.

Random Forests Algorithm can be performed by applying the following steps [37]:

Algorithm I: Random Forests Algorithm

Step 1:
Draw N tree bootstrap samples from the original data.

Step2:
For each of the bootstrap samples, grow an un-pruned classification or regression tree.

Step 3:
At each internal node, rather than choosing the best split among all predictors, randomly select m try of the M predictors and determine the best split using only those predictors.

Step 4:
Save tree as is, alongside those built thus far (Do not perform cost complexity pruning).

Step 5:
Predict new data by aggregating the predictions of the N trees.

The predictions of the Random Forests are taken to be the majority votes of the predictions of all trees for classification and for regression are taken to be the average of the predictions of the all trees as shown in "equation (1)" [37]-[39]:

$$S = \frac{1}{K}\sum_{k=1}^{k} K^{th} \quad (1)$$

Where S is a random forests prediction, K^{th} is a tree response, and K is the index runs over the individual trees in the forest.

The random forest error rate depends on two things:

1) Correlation: represents correlation between any two trees in the forest. Error increases as the correlation increases.
2) Strength: represents the strength of each tree in the forest.
The strength is measured by the error rate; a tree with low error is a strong tree. The forest error rate decreases as the decision tree's strength increases.

One of the advantages of random forest classifier is that it is one of the highly accurate classifiers. On the other hand, it has been observed to over-fit for some datasets with noisy classification tasks.

IV. FEARTURE SELECTION

Feature selection (FS) is an important pre-processing step to identify the important features and removing irrelevant (redundant) ones from the dataset and so reduce feature dimensions for classification. Generally the feature selection objectives are data dimensionality reduction, improving prediction performance, and good data understanding for different machine learning applications [38]. Feature selection is mandatory due to the abundance of noisy, irrelevant, or misleading features. The selected features will improve the performance of the prediction model and will provide a faster and more cost effective prediction than using all the features. FS can be seen as a combinatorial optimization problem that involves searching the space of possible feature subsets to identify the optimal (best) feature space separability, where the classification error is the function to be minimized [40], classification accuracy or some other criterion that might consider the best trade-off between attributes. Previously, an exhaustive search for the optimal set of features (attributes) in a high dimensional space may be unpractical [41]-[42].

Feature selection can be divided into four categories; Filter method is independent from learning method and uses measurement techniques such as correlation and distance measurement to find a good subset from entire set of features. Wrapper method uses pre-determined learning algorithm to evaluate selected feature subsets that are optimum for the learning process. Hybrid method combines advantage of both Filter and Wrapper method together. It evaluates features by using an independent measure to find the best subset and then uses a learning algorithm to find the final best subset. Finally, embedded method interacts with learning algorithm but it is more efficient than Wrapper method because the filter algorithm has been built with the classifier.

In search space the size is exceeds exponentially with respect to the number of attributes in the data set used, so in practice the exhaustive search is impossible in almost cases. A diversity of search technique has been utilized to solve the FS problem, such as greedy search based on sequential forward selection (SFS) and sequential backward selection (SBS). However, these attribute reduction approaches still suffer from several of issues, such as stagnation in local optima and increasing in the cost of computational. So as to improve the attribute reduction issues, an efficient global search algorithm is needed. Evolutionary computation (EC) algorithms are well-known for their global search capability. Grey wolf optimization (GWO) is a comparatively recent EC algorithm, that is computationally less expensive than some another EC techniques.

A. Grey Wolf Optimization

Grey wolf optimization is illustrated briefly in the following subsections based on the research work in [9]-[44].

1) Inspiration

Grey wolves are species with very strict social dominant hierarchy of leadership. The leaders are a male and a female, called alpha. The alpha is mostly responsible for making decisions about hunting, sleeping place, time to wake, and so on. The alphas decisions are dictated to the pack.

The second level in the hierarchy of grey wolves is beta. The betas are subordinate wolves that help the alpha in decision-making or other pack activities. The beta wolf is the best candidate to be the alpha in case one of the alpha wolves passes away or becomes very old to lead. The lowest ranking grey wolf is omega. The omega plays the role of scapegoat. Omega wolves always have to submit to all the other dominant wolves. They are the last wolves that are allowed to eat. The fourth class is called subordinate (or delta in some references). Delta wolves have to submit to alphas and betas, but they dominate the omega. *Scouts,* *sentinels, elders, hunters*, and *caretakers* belong to the delta category and each has its own defined responsibilities.

2) Mathematical Modelling

The GWO the fittest solution is called the alpha (α)while the second and third best solutions are named beta (β) and delta (δ) respectively. The rest of the candidate solutions are assumed to be omega (ω). The hunting is guided by α, β, and δ and the ω follow these three candidates. In order for the pack to hunt a prey they first encircling it. In order to mathematically model encircling behavior the following equations are used.

$$\vec{X}(t+1)=\overrightarrow{X_P}(t)+\vec{A}.\vec{D} \quad (2)$$

Where $\vec{D}$ is defined in 3 and t is the number of iteration,$\vec{A}$, $\vec{C}$, are coefficient vectors, $\vec{X}_P$ is the prey position and $\vec{X}$ is the grey wolf position.

$$\vec{D}=\left|\vec{C}.\overrightarrow{X_P}(t)-\vec{X}(t)\right| \quad (3)$$

The $\vec{A}$, $\vec{C}$ vectors are calculated as in "equation 4" and "equation 5" as follow:

$$\vec{A}=2\vec{A}.\vec{r}_1-\vec{a} \quad (4)$$

$$\vec{C}=2.\vec{r}_2 \quad (5)$$

Where components of $\vec{a}$ are linearly decreased from 2 to 0 over the course of iterations and r_1, r_2 are random vectors in[0,1]. The hunt is usually guided by the alpha. The beta and delta might also participate in hunting occasionally. In order to mathematically simulate the hunting behavior of grey wolves, the alpha (best candidate solution) beta, and delta are assumed to have better knowledge about the potential location of prey. The first three best solutions obtained so far and oblige the other search agents (including the omegas) to update their positions according to the position of the best search agents. So the updating for the wolves positions is as in "equations 6", "equation 7", and "equation8"

$$\overrightarrow{D_\alpha}=\left|\overrightarrow{C_1}.\overrightarrow{X_\alpha}-\vec{X}\right|,\overrightarrow{D_\beta}=\left|\overrightarrow{C_2}.\overrightarrow{X_\beta}-\vec{X}\right|,\overrightarrow{D_\delta}=\left|\overrightarrow{C_3}.\overrightarrow{X_\delta}-\vec{X}\right| \quad (6)$$

$$\overrightarrow{X_1}=\left|\overrightarrow{X_\alpha}-\overrightarrow{A_1}.\overrightarrow{D_\alpha}\right|,\overrightarrow{X_2}=\left|\overrightarrow{X_\beta}-\overrightarrow{A_2}.\overrightarrow{D_\beta}\right|,\overrightarrow{X_3}=\left|\overrightarrow{X_\delta}-\overrightarrow{A_3}.\overrightarrow{D_\delta}\right| \quad (7)$$

$$\vec{X}_{(t+1)}=\frac{\overrightarrow{X_1}+\overrightarrow{X_2}+\overrightarrow{X_3}}{3} \quad (8)$$

An important note about the GWO is the updating of the parameter a that controls the tradeoff between exploitation and exploration. The parameter $\vec{a}$ is linearly updated in each iteration to range from 2 to 0 according to the "equation 9".

$$\vec{a}=2-t.\frac{2}{Max_{iter}} \quad (9)$$

Where t is the iteration number and Max_{iter} is the total number of iteration allowed for the optimization.

Algorithm 2: GWO Search Algorithm

Input: N number of wolves (agents) used
NIter number of iterations for optimization.
Output: X_α Optimal wolf position
$f(X_\alpha)$ Best fitness value

1) Initialize a population of N wolves' positions at random,
2) Find $\alpha, \beta,$ and δ solutions based on their fitness values
3) Calculate the $\vec{a}$ parameter given the current iteration and the maximum number of iterations using "equation 9"
4) **While** Stopping criteria not met **do**

for each $\boldsymbol{Wolf_i}$ **do**
Update the current $\boldsymbol{Wolf_i}$ position according to "equation 8"
end
I. Update a, A, C
II. Evaluate the positions of individual wolves
III. Update α , β, and δ
End

V. THE PROPOSED HYBRID PREDICTION (CLASSIFICATION) ALGORITHM

The proposed ensemble classification algorithm consists of different phases as explained in (Fig.1)

Algorithm 3: The Proposed Ensemble Classification Algorithm

1) Data Pre-Processing,
2) Apply the Grey Wolf Optimizer,
3) Feature extraction & selection (Apply the Wrapper Approach**),**
4) Classification (Apply the Random Forests method**),**
5) Stopping Criterion (If Maximum No. of iterations > *NIter*, then go to Step 6; otherwise go to step2).
6) Results & performance analysis.

A. Data Pre-Processing Phase

The data used in our suggested prediction system is real world data about terrorist attacks occurred in Egypt along the period from 2006 till 2014 from the global terrorism database (GTD). The data are required to be prepared for using in the classification process and it passed on multiple steps as explained below:

1) Convert data from text format into categorical data format.
2) The features in our data are divided into 3 different types (Time domain features, Position domain features, Attack type features)
3) Calculate the correlation between the data features (attributes, predictors) and the class (Response) attribute.
4) Determine & Select the most relevant features to the class (Response) attribute.
5) Due to the huge number of attributes (features) in our data; we had to apply a K-Means clustering method in order to minimize the total size of the data attributes.
6) We transformed our used data from categorical form into binary data format to be numeric; in our study we based on applying M-Category attribute approach by using XLMiner.

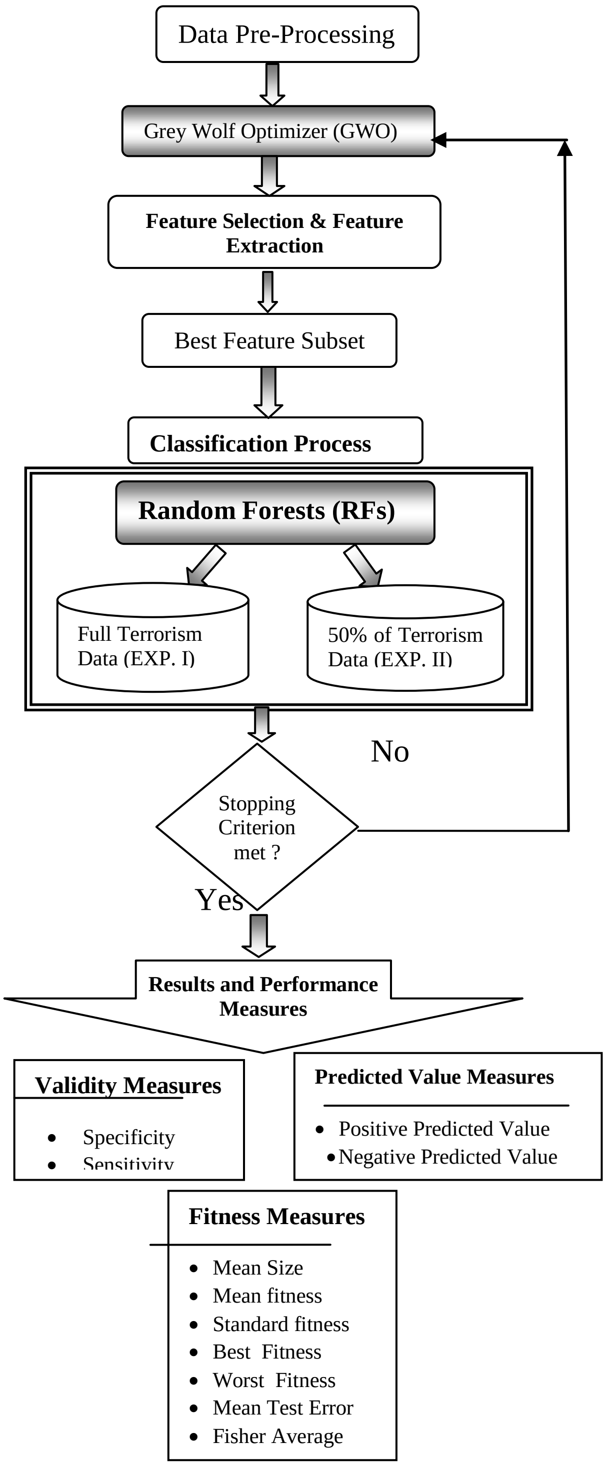

Fig. 1 Proposed Hybrid Algorithm Framework

B. Feature Selection & Extraction Phase

A wrapper approach for feature selection and attribute reduction is used in our study; where the attribute space that consists from 51 attributes is explored to find an attribute (feature) subset guided by classification performance of individual attribute subsets. Hence intelligent exploration of search space is always a challenge as the single evaluation of fitness function is always time consuming. This approach may be slow since the optimizer (GWO) must be retrained on all candidate subsets of the attribute set and its performance must be also measured to find the attribute combination that maximizes the following fitness function.

$$\text{Fitness} = CCR(D) \qquad (10)$$

Where CCR(D) is the correct classification ratio at feature set *D*. On the other hand wrapper approach searches a very large space of attribute combinations which it may be inefficient but it is much classifier guided and hence; if efficiently used, it can has a better performance.

The used fitness function in “equation 10” represents the predictability of attributes from each other and the predictability between individual features. Hence the goodness of an attribute combination is estimated as how much the selected attributes can correctly predict the output class labels and how much are they dependent. The convergence speed for GWO is ensured for its efficient searching capability and for the simplicity of the used fitness function. This step of optimization is stopped at a predetermined number of iterations as explained in Algorithm3.

C. Classification Process Phase

The data used about terrorism is divided into 3 equal parts; one for training the classifier, the second for validation and the third for testing the model.

GWO algorithm results’ are compared with particle swarm optimization (PSO) and Genetic Algorithm (GA) as they are known with their popularity in space searching. The classification process of the terrorist groups of attacks is performed based on RF ensemble classifier which compared with KNN classifier. A simple and commonly utilized learning algorithm [37], KNN is utilized in the experiments based on trial and error basis where the best choice of (K=5) is selected.

Through the training process, every wolf position represents one attribute subset. Training set is used to evaluate the RF ensemble classifier which is compared with KNN classifier; on the validation set throughout the optimization to guide the feature selection process. The test data are kept hidden from the optimization and is left for final evaluation.

The global and optimizer-specific parameter setting is outlined in Table I. All the parameters are set either according to domains specific-knowledge as the α; β parameters of the used fitness function, or based on trial and error on small simulations and common in the literature such as the rest of parameters.

TABLE I
Parameter Setting for Experiments

Parameter	Value
No. of search agents	8
No. of iterations	70
Problem Dimension	51
Search Domain	The given data set of terrorism
No. of Repetition of Runs	20
Inertia Factor of PSO	0.1
Individual Best Acceleration of PSO	0.1
Crossover Fraction in GA	0.8
α Parameter in the fitness Function	0.99
β Parameter in the fitness Function	0.01

VI. EXPERIMENTAL RESULTS AND ANALYSIS

The experiments are conducted on the terrorism data over two different trials; one during the whole data set where the search domain is 740 instances (data records) and 51 features, and the second experimental trial conducted over 50% of the whole data as illustrated in the following tables:

TABLE II
Fitness Results of the Classification Process by Different Classifiers and Various Optimizers applied on full data used (EXP. I).

Fitness Value	GWO	GA	PSO
RF **KNN**	**0.38** **0.41**	0.48 0.46	0.43 0.44
RF **KNN**	**0.39** 0.41	0.44 **0.40**	0.40 0.43
RF **KNN**	0.32 **0.33**	0.36 0.36	**0.30** 0.38
RF **KNN**	**0.37** **0.38**	0.42 0.40	0.39 0.43
RF **KNN**	0.38 **0.38**	0.40 0.43	**0.37** 0.41
RF **KNN**	**0.31** **0.36**	0.38 0.37	0.35 0.36
RF **KNN**	**0.36** **0.38**	0.44 0.41	0.38 0.40
RF **KNN**	**0.39** 0.41	0.40 **0.38**	0.38 0.39
RF **KNN**	0.38 **0.32**	0.37 0.42	**0.35** 0.41
TOTAL	**6.66**	7.32	7

Table II and table III summarize the results of running the different optimization algorithms for 20 runs by RF and KNN classifiers.

Fitness value obtained by GWO achieves remarkable advance over PSO and GA among the two experiments which ensures the searching capability of GWO.

Fig.2 and Fig. 3 show how the GWO is effective in the fitness values and hence in the classification accuracy than GA, and PSO in both Experiments, and also outline that RF ensemble classifier performs competitively with KNN classifier.

TABLE III
Fitness Results from the Classification Process by Different Classifiers and Various Optimizers applied on 50% of the data used (EXP.II).

Fitness Value	GWO	GA	PSO
RF **KNN**	**0.17** **0.19**	0.26 0.24	0.19 0.28
RF **KNN**	**0.16** **0.19**	0.22 0.18	0.18 0.18
RF **KNN**	**0.20** **0.23**	0.37 0.23	0.23 0.28
RF **KNN**	**0.24** **0.21**	0.29 0.23	0.25 0.28
RF **KNN**	**0.20** **0.22**	0.28 0.23	0.20 0.28
RF **KNN**	**0.17** **0.18**	0.20 0.19	0.18 0.23
RF **KNN**	**0.17** **0.18**	0.28 0.19	0.19 0.21
RF **KNN**	**0.21** 0.21	0.30 **0.20**	0.21 0.24
RF **KNN**	**0.18** **0.20**	0.24 0.24	0.20 0.27
TOTAL	**3.51**	4.37	4.15

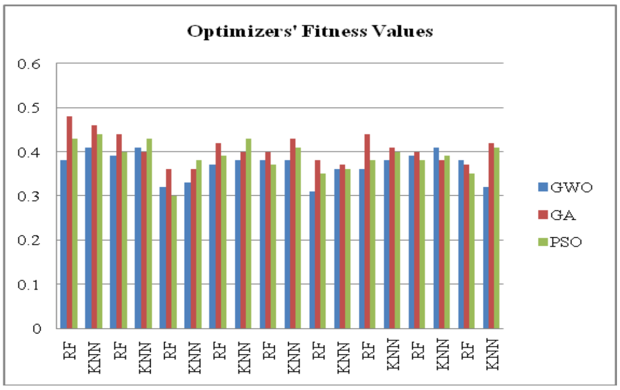

Fig. 2 Fitness Value for each Classifier by the Optimizers used from (EXP. I)

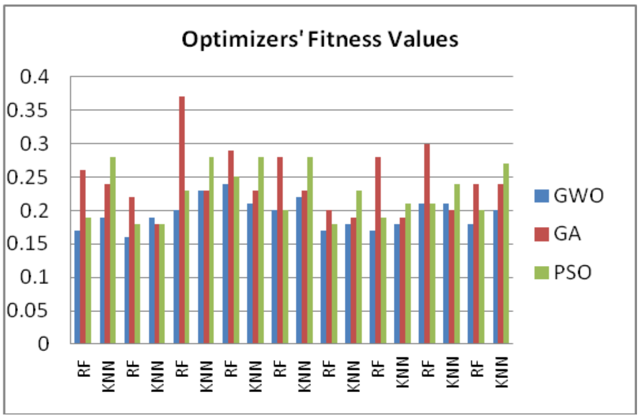

Fig. 3 Fitness Value for each Classifier by the Optimizers used from (EXP. II)

TABLE IV
Evaluation Criteria Results' of different Optimizers by EXP.I

Evaluation Criteria	GWO	GA	PSO
Mean Fitness	**0.376112**	0.403054	0.406134
Std. Fitness	**0.029177**	0.0317965	0.037072
Best Fitness	**0.321118**	0.363269	0.361784
Worst Fitness	**0.414609**	0.463404	0.439016

Table IV and Table V outline the fitness performance of different optimizers conducted from multiple experiments; where the GWO shows high fitness performance over the GA, and PSO algorithms in which it has the lowest mean fitness and as well as has lowest standard deviation of the obtained fitness values that proves the optimizer stability, repeatability of convergence and robustness.

TABLE V
Evaluation Criteria Results' of different Optimizers by EXP.II

Evaluation Criteria	GWO	GA	PSO
Mean Fitness	**0.201873**	0.213717	0.250079
Std. Fitness	**0.017519**	0.024527	0.039268
Best Fitness	**0.175509**	0.175589	0.175587
Worst Fitness	**0.228997**	0.240340	0.284623

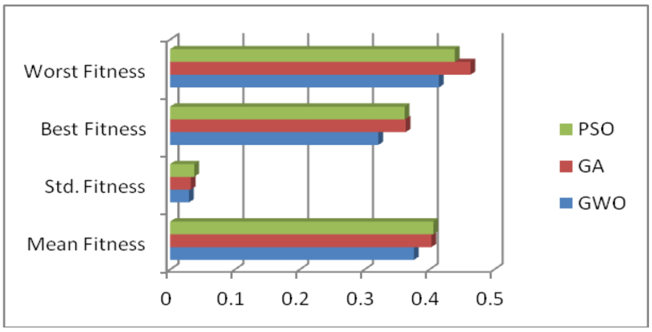

Fig.4 Fitness Measures' Results of GWO, GA, and PSO Optimizers By EXP. I

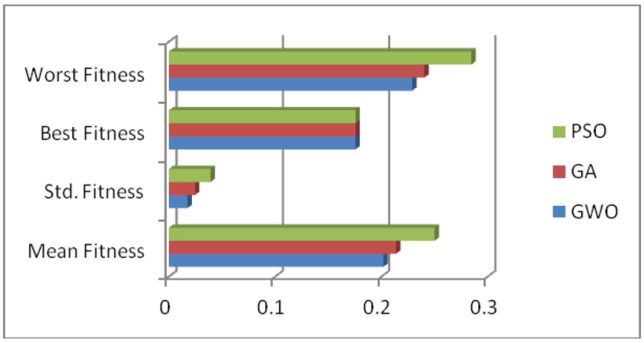

Fig.5 Fitness Measures' Results of GWO, GA, and PSO Optimizers 4RBY EXP. II

Fig. 4 and Fig.5 present an obvious view about the different fitness results for the GWO and other optimizers in both experiments where we can notice that GWO has the lowest and efficient results above the other optimizers which prove its capability and efficiency than GA, and PSO algorithms in the search space.

Table VI outlines the measures of validity of the used optimizers which measured by the sensitivity and specificity of a test; where we can conclude the superiority of GWO over GA, and PSO algorithms especially with RF classifier than KNN.

TABLE VI

Sensitivity & Specificity measure results of the experiments for the optimizers via different classifiers

Data set		*Full D*	*Data*		*Half*	*Data*	
		GWO	**GA**	**PSO**	***GWO***	***GA***	**PSO**
Sensit - ivity	**KNN**	0.4647	0.4431	**0.4714**	0.6705	0.6657	**0.6912**
	RFs	**0.4661**	0.4058	0.4409	**0.7001**	0.5780	0.5251
Speci- ficity	**KNN**	**0.9443**	0.9309	0.9294	**0.9641**	0.9374	0.9577
	RFs	**0.9249**	0.9170	0.9083	**0.9392**	0.9205	0.9055

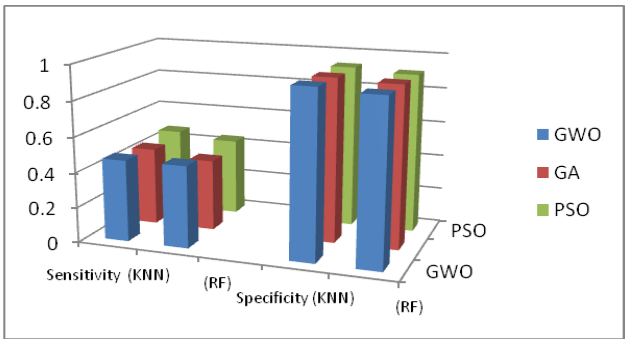

Fig. 6 Sensitivity and Specificity Results for the GWO, GA, and PSO by (EXP.I)

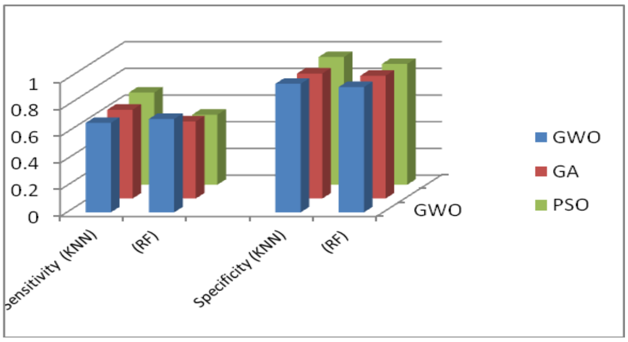

Fig. 7 Sensitivity and Specificity Results for the GWO, GA, and PSO by (EXP.II)

Fig. 6 and Fig. 7 outline and show the sensitivity and specificity results that used as measures for the validity of the algorithm where we can conclude that GWO has the highest results above GA, and PSO algorithms, the figures show also the competitive result of RF classifier with respect to KNN algorithm.

VII. CONCLUSIONS AND FUTURE WORK

The paper proposed a hybrid ensemble classification algorithm based on combining GWO and RF with the help of Wrapper feature selection approach that can be used in the prediction of terrorist groups among different regions and countries. The proposed model implements grey wolf optimizer (GWO) and wrapper feature selection approach in order to select optimal feature subset for classification process based on random forests (RFs) ensemble classifier to improve and enhance the classification accuracy while minimizing the number of selected features. The performance of the hybrid GWO-RFs model is tested by two different experiments during 20 iterations and the results are benchmarked for evaluation with particle swarm optimization (PSO) and genetic algorithm (GA), as well as the results of RF classifier are compared with another well known classifier as K-nearest neighbor (KNN). A set of assessment indicators are used to evaluate and compare between the obtained results which prove the capability of the proposed hybrid GWO-RFs algorithm to search the feature space for the optimal feature combination as well as enhancing the classification accuracy compared to other well-known conventional, heuristics and meta-heuristics search algorithms. Experimental results demonstrate competitive performance of the Hybrid GWO-RF ensemble classification algorithm, especially with high dimension datasets.

Further investigation on the parameters values and testing the proposed hybrid GWO-RF algorithm with other feature selection approaches on different dimensions data sets are different and various areas for future research.

REFERENCES

[1] A. E. Eiben, et al.,"Genetic algorithms with multi-parent recombination", PPSN III: Proceedings of the International Conference on Evolutionary Computation. *The Third Conference on Parallel Problem Solving from Nature;* 1994: 540-584.

[2] M. Dorigo, M. Birattari, T. Stutzle, "Ant colony optimization", Comput *Intell Magaz,IEEE* 2006;1:28–39.

[3] J. Kennedy, R. Eberhart, "Particle swarm optimization, in Neural Networks", In: *Proceedings, IEEE international conference on*; 1995. p. 1942–1948.

[4] R. Storn, K. Price, "Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces". *J Global Optim* 1997;11:3 41–59.

[5] X. Yao, Y. Liu, G. Lin., " Evolutionary programming made faster", *Evolut Comput,IEEE Trans* 1999;3:82–102.

[6] N. Hansen, SD. Müller, R. Koumoutsako, ”Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMAES)”, Evolut Comput 2003;11: 1–18.

[7] D.H. Wolpert, W.G. Macready, No free lunch theorems for optimization, IEEE Trans. Evol. Comput.1 (1997) 67–82.

[8] S. Mirjalili, “Moth-flame Optimization algorithm: A novel nature-inspired heuristic paradigm. Knowledge-Based Systems”, 2015, 228-249.

[9] S. Mirjalili, L. Andrew, “Grey wolf- optimizer”, *Advances in Engineer Software* ,2014, 46-61.

[10] Garsava, Gintautas and Danenas, Paulius, “Particle swarm optimization for linear support vector machine based classifier selection”, *I Journal of Nonlinear analysis: Modeling and Control*. 2014; 26-42, vol.19. no. 1.

[11] S. M. Kamruzzaman, H. Farhana, and R. Ahmed. *Text Classification using Data mining*. arXiv preprint arxiv,2010, 1009.4987.

[12] B. G´erard, D. Luc, L. G´abor, “Consistency of Random Forests and Other Averaging Classifiers”, *Journal of Machine Learning Research* 9 ,2008, 2015-2033.

[13] M. H. Khorshid , H. M. Tarek Abou-El-Enien , M. A. Ghada Soliman,”A Comparison among support vector machine and other machine learning classification algorithms”. *IPASJ International Journal of Computer Science (IIJCS),* 2015Volume 3, Issue 5

[14] W. Siedlecki, J. Sklansky, “A note on genetic algorithms for large-scale feature selection”, *Pattern Recognition Letters,* 10(5) , 335-347, 1989.

[15] X. Wang, J. Yang, X. Teng, w. Xia, “ Feature selection based on rough sets and particle swarm”, *Pattern Recognition Letters*”, 28(4), 459-471, 2007.

[16] S. Parpinelli, S. Lopes,F. Alex,”Data mining with an Ant Colony Optimization Algorithms”, *IEEE Trans. Evol. Comput*, Vol. 6, no. 4, pp. 321-332, 2002.

[17] Hang, L. Cheng, “ACO-based hybrid classification system with feature selection and model parameters optimization*”, Neurocomputing* 73, 2009, 438- 448.

[18] H. Kanan, K. Faez, and T. Sayyed, *Feature selection using Ant Colony Optimization (ACO): a New Method and Comparative Study in the Application of Face Recognition System*, P Perner (ed.), ICDM 2007, LNAI 4597, pp. 63-76, 2002.

[19] S. Mohamed and H. Aboul Ella, ” Incorporating Random Forest Trees with Particle Swarm Optimization for Automatic Image Annotation”, Preprints *of the Federated Conference on Computer Science and Information Systems* pp. 791–797, 2012.

[20] A. Whitney, “A direct method of nonparametric measurement selection”. *IEEE Trans. Comput*; 1971; 20: 1100-1103.

[21] T. Marill and D. Green, “On the effectiveness of receptors in recognition systems”. *IEEE Trans. Inf. Theory*; 1963; 9: 11-17.

[22] S. Stearns, “ On selecting features for pattern classifier”. *Proceedings of the 3rd International Conference on Pattern Recognition, IEEE*, Coronado, CA; 2003: 7175.

[23] H. Almuallim, T.G. Dietterich, “Learning Boolean concepts in the presence of many irrelevant features. Artificial Intelligence”; 1994; 69: 279-305.

[24] L. Breiman, “Bagging predictors. Machine Learning”,1996,24: 123–140.

[25] Y. Freund and R. Schapire, *Experiments with a new boosting Algorithm.* In L. Saitta (Ed.), Machine Learning: Proceedings of the 13th International Conference, pp. 148–156, Morgan Kaufmann, San Francisco, 1996.

[26] L. Breiman, ”Arcing classifiers. The Annals of Statistics”,(1998) 24:801–849.

[27] T.G. Dietterich, “An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. Machine Learning”, 40:139–157, 2000.

[28] L. Breiman,” Random forests. Machine Learning”, 45:5–32, 2001.

[29] R. Duangsothong, T. Windeatt, ”Hybrid Correlation and Causal Feature Selection for Ensemble Classifiers”, 2011.

[30] L. Breiman. *Some infinite theory for predictor ensembles.* Technical Report 577, Statistics Department, UC Berkeley, 2000. http://www.stat.berkeley.edu/˜breiman .

[31] A. Cutler and G. Zhao. Pert , “Perfect random tree ensembles”, *Computing Science and Statistics,* 33:490–497, 2001.

[32] Y. Amit and D. Geman, “Shape quantization and recognition with randomized trees*”. Neural Computation*, 9:1545–1588, 1997.

[33] L. Devroye, L. Gy¨orfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition*. Springer-Verlag, New York, 1996.

[34] L. Breiman. *Consistency for a simple model of random forests.* Technical Report 670, Statistics, Department, UC Berkeley, 2004.

[35] Y. Lin and Y. Jeon, “ Random forests and adaptive nearest neighbors”. *Journal of the American Statistical Association*, 101:578–590, 2006.

[36] N. Meinshausen, “Quantile regression forests”. *Journal of Machine Learning Research*,7:983–999,2006.

[37] Y.K. Vrushali and K. S. Pradeep, “Efficient Learning of Random Forest Classifier using Disjoint Partitioning Approach”, *in Proceedings of the World Congress on Engineering*.Vol. 2. 2013.

[38] G. Chandrashekar, F.Sahin, “A survey on feature selection method”s, *Comput. Electr. Eng*. 40(1), 16–28,2014.

[39] B. Anna , Z. Andrew and M. Xavier, “ Image Classification using Random Forests and Ferns*”, IEEE 11th International Conference on Computer Vision*, 2007.

[40] M. H. Rasmy, M. El-beltagy, M. saleh, B. mostafa, “A Hybridized Approach for Feature Selection using Ant Colony Optimization and Ant-Miner for Classification”, *International Conference on INFOrmatics and Systems*(INFOS2016) -14-16 May.

[41] E. Emary, HossamM.Zawba, AboulEllaHassanien, “Binary grey wolf optimization approaches for feature selection”. *Neurocomputing*172 (2016)371–381.

[42] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification, seconded*, Wiley- Interscience Publication,USA,2000.

[43] E. Emary, Y.Waleed, H. Aboul Ella and S. Vaclav, “Multi-Objective Gray-Wolf Optimization for Attribute Reduction”, *International Conference on Communication, Management and Information Technology* (ICCMIT2015).

[44] L.Y. Chuang, H.W .Chang, C.J.Tu, C.H. Yang, ” Improved binary PSO for feature selection using gene expression data” ,*Comput.Biol.Chem*.32(2008)29–38.

Ghada M. A. Soliman received her degree in B. of science in Operations Research and Decision Support, at the faculty of Computers and Information, Cairo University in 2001. Then she received her Master Degree in Operations Research in 2007, she is currently a PhD student. Her employment experience includes FCI institution since 2001; she is employed with the Department of Operations Research as Lecturer Assistant. Her research interests include modeling, simulation, data mining and soft computing.

Prof. Dr. Motaz M. H. Khorshid, is a Professor of Operations Research, - Department of O.R. & D. S. - Faculty of Computers Information-Cairo University**.** His research interest is focused on computer aided Modeling and Decision Support Systems (DSS), Computer Simulation Methodology, Economy-wide Modeling and Analysis , Socioeconomic Data Systems, Planning and Management of Higher Education and Scientific Research , Development Planning, Quantitative Management and Corporate Planning Models .

Dr. Tarek H. M. Abou-El-Enien is an Associate Professor of Operations Research, - Department of O.R. & D. S. - Faculty of Computers & Information- Cairo University. His research interests in decision analysis, soft computing, modeling, data mining and multi-level multi-objective Programming field