Digitally Modulated Signal Recognition based on Feature Extraction Optimization and Random Forest Classifier

Abstract.. In the past decades, there was a growing need for the automatic classification that is related to digital signal formats, that also appears to be on going tendency in the future. Automatic modulation recognition (AMR) is con-sidered to be of high importance in military and civil applications and communication systems . The recognition regarding the received signal modulation can be defined as a transitional stage between detection and demodulation of signals.in this paper , several features which are associated with the received signal will be extracted and used. which is of high importance in increasing the AMR's effectiveness.algorithms from Chicken Swarm optimization and Bat Swarm optimization were used to improve the features of modulated signals and thus increase the accuracy of the classification. It then classifies the features of the modified signals resulting from the optimization algorithms by a random forest . The results showed thatswarm Chickens algorithm performs better than the Bat swarm algorithm, even at level SNR low, Results From chicken Algorithm &Random forest classifier were Accuracy of classification 95% while Accuracy of classification From Bat Algorithm &Random forest 91%.

Keywords: : Chicken Swarm Optimization algorithm, Random Forests classifier, Bat Swarm Optimization algorithm.

1. Introduction

AMR can be defined as a method which report the modulation types regarding the received signals in automatic way. Modulation recognition is very essential in communication intelligence (COMINT). In COMINT one needs to recognize the modulation scheme to demodulate the signal information content. Furthermore, classifying right modulation type is helpful to choose an appropriate type of jamming to transmit. Signal modulation recognition is an interdiscipline composed of signal extraction, signal processing, signal analysis, pattern recognition and so on. As the complex nature of the received signals, there are still many problems to be solved urgently-feature extraction and modulation classifier [1].

In [2] . In this study, the researchers selected suitable features for the modified input signal and used the neural network algorithm,. The presence of Gaussian noise -5 dB to 20 dB .The results showed that the use of the NN algorithm resulted in a significant increase in the precision of recognition of the type of modification .

In [3] Using modified artificial bee colony (MABC) with Enhanced Particle Swarm Optimization (EPSO) for predicting the infiltration detection. The algorithms will be combined for the purpose of achieving improved results of optimization and obtain accuracy of classification .

In[4] Use bilateral support vector machine (SVM) for the purpose of determining digital modulation approaches. Modulation types consist of: BASK, BFSK, BPSK, 4ASK, 4FSK, QPSK and 16-QAM. Simulation demonstrates the higher abilities regarding suggested features in the digitally modified signal classification Even at a low SNR level .

In [5] Artificial Neural Network (ANN) algorithm was utilized for classifying the modification. In the presented study, the system of AMC have the ability of distinguishing between USB, QPSK, LSB, 8PSK, and AM modulation. The rate of accuracy that is related to the presented system in modulation classification process without the use of nonlinear conversions is 65.5 percent on the signal quality of 10 dB. Therefore, the precision of AMC with the use of nonlinear transformations on incoming signal is about 88.8 percent on signal quality 10dB.

In [6] using the decision tree A dual-carrier bus system support has been trained on the features which have been extracted from high-order stacking tool. The results have indicated that the average accuracy of identifying formatting signals can be realized more than 94% when SNR is -5 dB .

in fact ,paper submitted by Almaspour & Moniri, 2016 [2] under title " Automatic modulation recognition and classification for digital modulated signals based on ANN algorithms " The closest research with our study. ,we have used Same types of modified signals test the performance of the Classification .We used (CSO, BA)algorithem and Random Forest Classifier To improve system performance. while the researchers selected suitable features for the modified input signal and used the neural network algorithm to classified the signal, better results were obtained.

The major aim of this study is optimizing the features of the modulated signals, thus reducing the signal characteristics and thereby increasing the accuracy of the system in detection and identification of the signal type, using two types of Swarm optimization algorithms and comparing their performance to detect the best in strengthening the features .In this work, Random forest is as classified. Also, one of the main issues is selecting the suitable feature set. In previous paper , usually are numerous Features used to classify the composition leading to Improved efficiency. So many features are not effective enough one of the main reasons leads to restrictions Most of the techniques recognize a digital signal The features are related [7] .The general structure of this study is as follows. After introduction, features extraction and algorithms of optimization are going to be reviewed in Section 2, the classified will be presented in Section 3. Section 4 Provide some simulation results, and lastly the conclusions of the study will be presented in section 5.

2. Feature Extraction

The initial stage in extracting features is to model the modified digital signals for extracting the signal features. For the purpose of analyzing the implementation regarding the metering system [2] Various digital signal types have a lot of features. So find appropriate features to recognize the scheme of digitally modulated signals are a serious problem. The main features of the configuration classification should be identified in the pattern recognition approach. These features must be Strong properties are sensitive to modulation types and are not sensitive with SNR variance. Instantaneous .Features such as Moments and cumulants were calculated. These were used to classify signals Features [8]. k^th-order cumulant can be written in the following way based on the HOCs property

$$C_{k,n(x(t))} = C_{k,n(s(t))} + C_{k,n(n(t))} \qquad (1)$$

As the white Gaussian noise's value is considered to be larger compared to 2nd order cumulant constant to 0, i.e.

$$C_{k,n\,(n(t))} = 0, K > 2 \qquad (2)$$

Thus, the non-Gaussian signal that has been received and consist of white Gauss noise is going to be converted to cumulant processing, and thus the impact of noise will be removed. Typically, HOC will be defined as function related to the high-order moments (HOMs) of the signal. For complex k^th-order stationary stochastic procedure S(t) with zero average value, which its p^thh-order mixing moment [9] is

$$M_{pq=E[S\ (t)^{p-q} S^{*}(t)^{q}]} \qquad (3)$$

In which p will be referred to as the moment order and [.]^*represents complex conjugation. Symbolism for k^th-order of cumulant is comparable to that of pth-order moment. In particular:

$$C_{pq} = Cum\left[\underbrace{s(t), \ldots, s(t),}_{p-q} \underbrace{S^{*}(t), \ldots, S^{*}(t),}_{q}\right] \qquad (4)$$

Thus, depending on Eq. (3) and eq. (4), one might define an association between HOMs and HOCs of 2nd to 8th order cumulants are indicated in (Table1).

Table 1. The correlation between the HOCs and HOMs.

HOC_S	HOM_S Expresion
Second order Cumulant	$C_{2,0}$ $M_{2,0}$ $C_{2,1}$ M_{21}
Fourth order Cumulants	$C_{4,0}$ $M_{4,0} - 3M_{2,0}{}^{2}$
	$C_{4,1}$ $M_{4,1} - 3M_{4,0}, M_{21}$ $M_{4,1} - \lvert M_{21} \rvert^{2} - 2M_{2,1}{}^{2}$
Sixth order Cumulant	$C_{6,0}$ $M_{6,1}$ -15$M_{4,0}$ $M_{2,0}$ $M_{4,1}$ + 30 $M_{2,1}{}^{3}$
	$C_{6,3}$ $M_{6,3}$ -$M_{4,1}$ $M_{2,1}$ + 12$M_{2,1}{}^{3}$ − 3 $M_{2,0}$ $M_{4,3}$ − 3$M_{2,2}$ $M_{4,1}$18$M_{2,0}$ $M_{1,2}$ $M_{2,2}$
Eight **order** **Cu**mulant	$C_{8,0}$ $M_{6,3}$ -28$M_{6,0}$ $M_{2,0}$ − 35$M_{4,0}{}^{2}$ + 420$M_{4,0}$ $M_{2,0}{}^{2}$ −630 $M_{2,0}{}^{4}$

1.1 Chicken swarm algorithm:

Chicken Swarm Optimization (CSO) can be defined as a novel intelligent bionic algorithm suggested based on different behaviors of chicks, hens, and cocks in their search for food. In CSO the chicken swarm in the search space will be mapped as particular particle individual. Chicken particle swarm, hen particle swarm, and cock particle swarm undergo sorting depending on the particle's fitness value, different search mode is utilized by each sub swarm [10] In CSO a lot of particles with optimum fitness will be chosen as the cock particle swarm, that is specified via

$$X_{i,j}^{t+1} = X_{i,j}^{t} + randn(0, \sigma^2)\,.X_{i,j}^{t} \qquad (5)$$

In which $X_{i,j}^{t+1}$ and $X_{i,j}^{t}$representing the location of the j^{th}dimension of the i-th particle in t+1 and t iterations respectively and randn (0, σ^2) can be defined as arbitrary number of the Gauss distribution whose variance = σ^2.The parameter σ^2might be estimated through

$$\sigma^2 = \begin{cases} 1, & fit_i < fit_k \\ \exp\left(\dfrac{(fit_k - fit_i)}{(|fit_i| + \xi)}\right), & fit_i \geq fit_k \end{cases} \qquad (6)$$

In which i, k∈ [1, r size] and $i \neq k$. r size are representing the number of the cock swarms. 〖fit〗_i and 〖fit〗_k are representing the fitness values regarding the cock particle i and k respectively; ξ representing the number that is small enough .Furthermore, the majority of particles that have optimum fitness have been chosen as the hen swarms. Its random search is achieved through cocks of hen population and that of others, that could be specified as

$$X_{i,j}^{t+1} = X_{i,j}^{t} + s2rand.\left(X_{r1,j}^{t}, X_{i,j}^{t}\right) + .s2\ rand.\left(X_{r2,j}^{t}, X_{i,j}^{t}\right) \qquad (7)$$

In which X_(r1,j)^t and X_(r1,j)^t are representing the position regarding the cock individual $r1$ in hen population xi and cock individual $r2$ in other population, respectively. rand is uniform random number over [0, 1]. $S1$ and $S2$ refer to the weight estimated via s1=$exp\left(\frac{(fit_{i-fit_{r1}})}{(|fit_i|+\xi)}\right)$

S2= $exp(fit_{r2} - fit_i)$ (8)

In which fit_i and $fit_k$are, respectively, cock's individual fitness value $r1$ in population of the hen xi and cock individual $r2$ in other population. Every individual, apart from hen swarm and cock swarm, are specified as chick swarm. Its mode of search mode follow that of the hen swarm, that is specified via

$$X_{i,j}^{t+1} = X_{i,j}^{t} + FL.\left(X_{m,j}^{t} - X_{i,j}^{t}\right), FL \in [0,2] \quad (9)$$

1.2 Bat swarm algorithm:

Bat Algorithm (BA) can be defined as swarm intelligence based algorithm, it depends on the echolocation behavior regarding the micro bats. When the bats are flying and hunting, they will be emitting certain ultrasonic, short, pulses to environment and list to their echoes. The studies have indicated that the information they obtain from the echoes are going to help the bat to construct accurate image regarding the environment around them and determining the shape, distance, and the location of the prey. This echolocation behavior of the bats is very important in finding the prey as well as distinguishing a lot of insects' types in total darkness. Previous studies indicated that the Bat Algorithm have the ability of solving un-constrained and constrained optimization problems with a lot more robustness and effectiveness in comparison to Genetic Algorithm and Particle Swarm Optimization[11] The following describes the execution steps of the typical BA:

Step 1: For each bat, initialize the position, velocity, and parameters and randomly generate the frequency.

Step 2: Update the position and velocity of each bat .

Step 3: For each bat, generate a random number (0< $rand_1$<1). Update the temp position and calculate the fitness value for corresponding bat if $rand_1$< r_i (*t*)

Step 4: For each bat, generate a random number (0< $rand_2$<1),, Update A_i (*t*) and r_i (*t*) if $rand_2$< A_i (*t*)and *f*(X_i (*t*))<*f*(*p*(*t*)).

Step 5: Sort each individual based on fitness values and save the best position

Step6: The algorithm is finished if the condition is met, otherwise, move on to Step 2 .

3. Classifier

3.1 Random forest

Random forest can be considered as ensemble classifier which contain a lot of decision trees and output the class which is mode of class's output via separate trees. The term was derived from the random decision forests which has been initially suggested via [12]. The approach combine feature’s random selection and the Breiman's "bagging" approach, presented. For the purpose of constructing set of decision trees with controlled variation. The random forests combine the tree predictors in a way that every one of the trees depend on the random vector values that are sampled independently and with same distribution for every forest tree. The generalization error regarding forest of the tree classifiers depend on strength regarding distinct trees in forest and association between them.

Random forests operate effectively on huge databases that can handle huge volumes of input variables with no deletion. It offers assessment regarding the significant variables in classification. Random forest in considered to be un-biased toward the assessment regarding the generalized error throughout forest formation. The algorithm also effectively calculate the missing data as well as preserving precision with approaches used to balance the errors in un-balanced class population datasets. The Resultant forests could be managed as inputs to future datasets. It provides information regarding the association between classification and variables. It operates extremely well in outlier detection, labeling of un-supervised clustering and data views Streaming Random Forest learning Algorithm .Random forest algorithm consist of the following steps:

Step1: Assuming that S is the number of training samples, whereas P is the number of variables in the classifier.

Step2: assume that p is the number of input variables that are utilized for determining decision at a node of the tree where p must be considerably smaller than P.

Step3: Choose a training set for certain tree via choosing S times with replacement from every S training sample available. Via predicting the classes, the remaining samples are utilized for estimating the error of the tree.

Step4: in order to make a decision at one of the nodes, randomly choose p variables for every one of the tree nodes. Calculate the optimal split in the training dataset according to the p variables.

Step5: every one of the trees will be grown to its maximum possible level so that there isn't any more pruning. [13] (Fig. 1) shows The structure of Random Forest

.

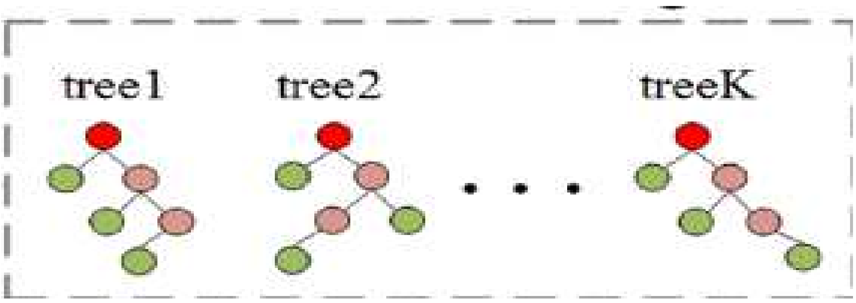

Fig1. The structure of Random Forest [14]

4. Proposed system

Was used Nine types (QAM8, QAM16, QAM32, QAM 64 , QAM 128,QAM 256, PSK2,PSK4, ASK2)of Modulated signals Which we have generated in the MATLAB Program within a level of SNR ranging from(-2,-1,0,1,2,3,4, 5) dB. After extracting the generated signal features. The extracted features were collected in a data matrix to improved it .This data Go through two stages : first phase is starting with optimization signal using two algorithms (CSO,BA). The second phase: is that takes the output from phase 1 as an input to random forest the used as a classifier to determine the accuracy of the classification and prediction of the type of signal. (Fig. 2) shows the proposed system. Where the algorithms used in the proposed system are mentioned in Section 3.

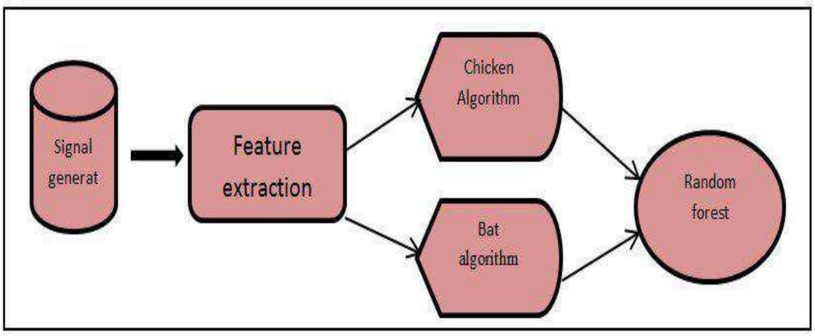

Fig 2. The schematic of the proposed approach

5. Experimental Result:

In this section, the results that have been obtained after Optimization the signal fea-tures modulated by the algorithm (Chicken& Random forest) and the results obtained after optimizing the features using a Java (Bat & Random forest) algorithms. The results of both techniques were compared.

1. The results obtained after execution of the SNR (-2,-1,0,1,2,3,4, 5) dB in Random Forest classification algorithm and Chicken algorithm are listed in the table2.

Table(2).represents results Accuracy Detailed of the chicken Algorithm

Type of signal	**Classification criteria**		
	Accuracy of **classification**	**Recall**	**F-Measure**
8 QAM	**%100**	**100**	**100**
16 QAM	**%100**	**100**	**100**
32 QAM	**%100**	**100**	**100**
64 QAM	**%100**	**100**	**100**
128 QAM	**%100**	**100**	**100**
256 QAM	**%100**	**100**	**100**
2 ASK	**%100**	**100**	**100**
2 PSK	**%100**	**100**	**100**
4 PSK	**%54**	**100**	**100**

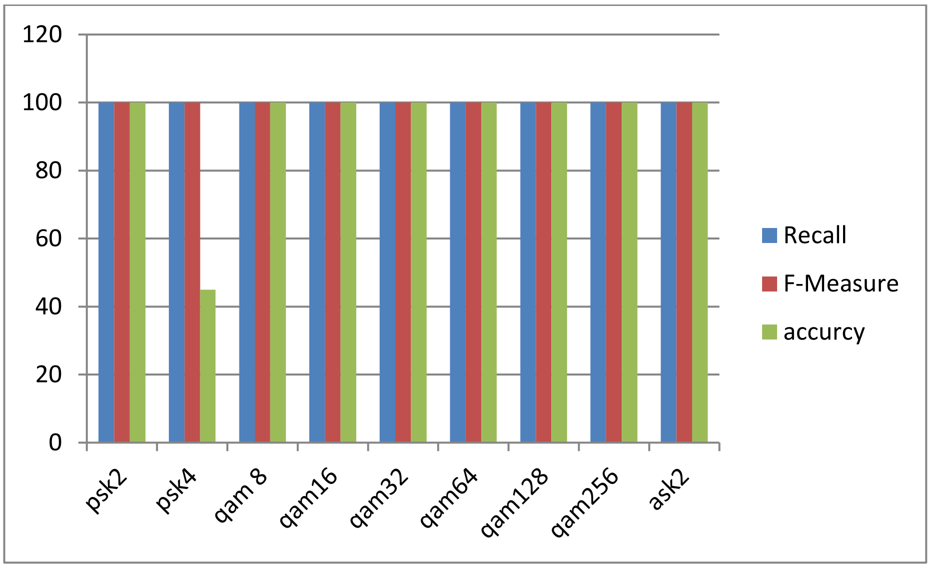

Fig. 3. represent the result obtain after optimization feature by CSO

2- The results obtained after execution of the SNR (-2,-1,0,1,2,3,4, 5) dB in Random Forest classification algorithm and Chicken algorithm are listed in the table3.

.

Table(3).represents results Accuracy Detailed of the Bat Algorithm

Type of signal	**Classification criteria**		
	Accuracy of classification	**Recall**	**F Measure**
8QAM	**%100**	**100**	**100**
16QAM	**%100**	**100**	**100**
32QAM	**%100**	**100**	**100**
64QAM	**%100**	**100**	**100**
128QAM	**%57**	**50**	**63**
256 QAM	**%100**	**100**	**100**
2 ASK	**%100**	**100**	**100**
4PSK	**%66**	**75**	**70**
2 PSK	**%100**	**87**	**93**

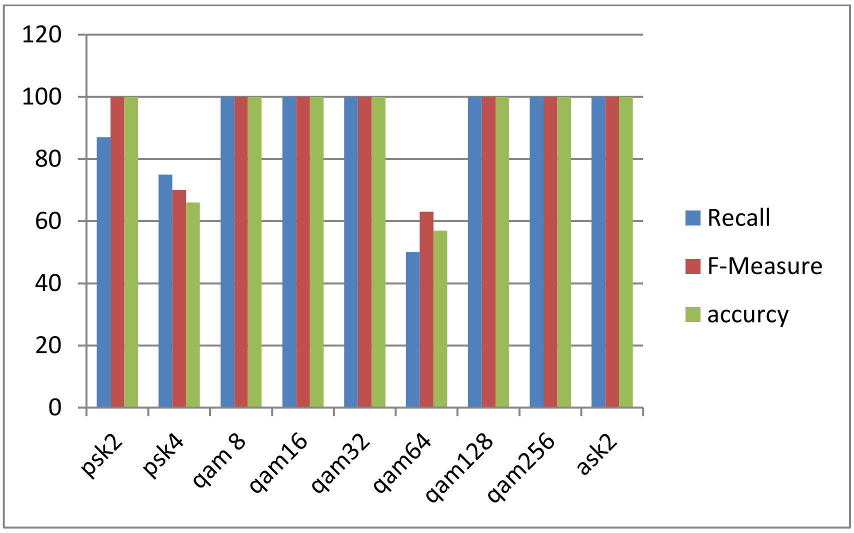

Fig 4. represent the result obtain after optimization feature by Bat algorithm

3-Comparison Between Performance (Chicken, Bat):

The results drawn from Two different techniques (chicken, Bat) optimization algorithm) and classification by Random Forest . Show chicken algorithm more efficiently than Bat algorithm has higher performance and higher classification accuracy .Been focusing on three basic criteria(precision, Recall, F-Measure.

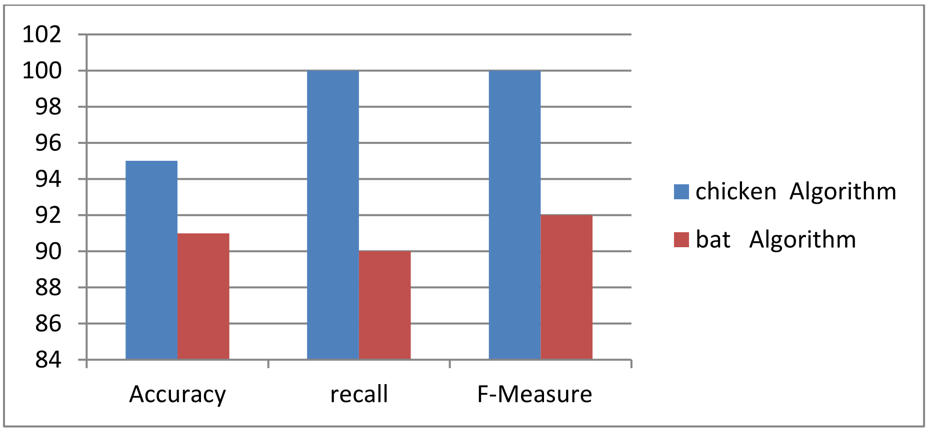

Fig 5. Comparison *between*(Bat &CSO)

6. Conclusion:

In this paper, Nine types of electromagnetic signals embedded in the MATLAB program were created within an SNR level ranging from (-2,-1,0,1,2,3,4,5)dB. Then we extract the statistical characteristics (moment, cumulant) of the signals, above. We improve the features, by two way Chicken swarm optimization algorithm and Bat Swarm Optimization algorithm. Using the Random Forest as a classifier for both. The results were compared of the proposed methods, the chicken swarm algorithm outperforms Bat Swarm algorithm by predicting the signal type and obtaining the highest classification accuracy even at a low SNR level of about 95%.

References

1. Cheng, L., Liu, J.: An optimized neural network classifier for automatic modulator recognition. TELKOMNIKA Indones. J. Electr. Eng. 12, 1343–1352 (2014).
2. Almaspour, S., Moniri, M.R.: Automatic modulation recognition and classification for digital modulated signals based on ANN algorithms. nine. 3, (2016).
3. Amudha, P., Karthik, S., Sivakumari, S.: A hybrid swarm intelligence algorithm for intrusion detection using significant features. Sci. World J. 2015, (2015).
4. Hassanpour, S., Pezeshk, A.M., Behnia, F.: Automatic digital modulation recognition based on novel features and support vector machine. In: 2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS). pp. 172–177. IEEE (2016).
5. Kurniansyah, H., Wijanto, H., Suratman, F.Y.: Automatic Modulation Detection Using Non-Linear Transformation Data Extraction And Neural Network Classification. In: 2018 International Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC). pp. 213–216. IEEE (2018).
6. Sun, X., Su, S., Huang, Z., Zuo, Z., Guo, X., Wei, J.: Blind modulation format identification using decision tree twin support vector machine in optical communication system. Opt. Commun. 438, 67–77 (2019).
7. Hakimi, S., Ebrahimzadeh, A.: Digital Modulation Classification Using the

Bees Algorithm and Probabilistic Neural Network Based on Higher Order Statistics. Int. J. Inf. Commun. Technol. Res. 7, 1–15 (2015).

8. Bagga, J., Tripathi, N.: Automatic modulation classification using statistical features in fading environment. Int. J. Adv. Res. Electr. Electron. Instrum. Eng. 2, 3701–3709 (2013).
9. Mirarab, M.R., Sobhani, M.A.: Robust modulation classification for PSK/QAM/ASK using higher-order cumulants. In: 2007 6th International Conference on Information, Communications & Signal Processing. pp. 1–4. IEEE (2007).
10. Liang, S., Feng, T., Sun, G.: Sidelobe-level suppression for linear and circular antenna arrays via the cuckoo search–chicken swarm optimisation algorithm. IET Microwaves, Antennas Propag. 11, 209–218 (2017).
11. Chakri, A., Khelif, R., Benouaret, M., Yang, X.-S.: New directional bat algorithm for continuous optimization problems. Expert Syst. Appl. 69, 159–175 (2017).
12. Barandiaran, I.: The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. 20, (1998).
13. Kumar, A., Kaur, P., Sharma, P.: A Survey on Hoeffding Tree Stream Data Classification Algorithms. CPUH-Research J. 1, (2015).
14. Li, K., Yu, N., Li, P., Song, S., Wu, Y., Li, Y., Liu, M.: Multi-label spacecraft electrical signal classification method based on DBN and random forest. PLoS One. 12, e0176614 (2017).