IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received June 2, 2019, accepted June 30, 2019, date of publication July 3, 2019, date of current version July 24, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2926615

A Novel Hybrid Cuckoo Search- Extreme Learning Machine Approach for Modulation Classification

SYED IHTESHAM HUSSAIN SHAH[1], SHERAZ ALAM[1], SAJJAD A. GHAURI[2], ASAD HUSSAIN[1], AND FARAZ AHMED ANSARI[1]

[1]Department of Electrical Engineering, National University of Modern Languages, Islamabad 44000, Pakistan
[2]Department of Electrical Engineering, ISRA University, Islamabad, Pakistan

Corresponding author: Sheraz Alam (salam@numl.edu.pk)

ABSTRACT This paper presents a novel hybrid extreme learning machine (ELM) with cuckoo search algorithm (CSA) for the classification purposes of the digitally modulated signals, such as phase shift keying (PSK), frequency shift keying (FSK), and quadrature amplitude modulation (QAM). Nine modulation schemes having different orders have been considered for this paper. First, the Gabor filter is used to extract the key features from the received signal which are then optimized by the CSA. Finally, the ELM is used to classify the modulation schemes using these optimized features. Our proposed CSA-ELM approach is not only fast convergent and robust but also manifests improved percentage classification accuracy at low SNRs and lower sample size for both AWGN and Rayleigh fading channels.

INDEX TERMS Extreme learning machine, cuckoo search algorithm, Gabor Features, automatic modulation recognition.

I. INTRODUCTION

Automatic modulation classification (AMC) is a process of automatic detection of modulation format from a received signal with no prior information (carrier, signal power, phase offset), also termed as blind classification. AMC is an established research field having various military and civil applications. It has been explored vigorously by researchers to improve the classification accuracy using different methods. There are different AMC approaches in the literature that can be divided into two broad categories as decision-theoretic or likelihood-based (LB) approach and Feature-based (FB) approach. Research work presented in this article is based on FB approach where first the reference features are extracted from the received signal and the decision is made using the calculated features based on the theoretical reference values for different modulation parameters [1].

From a practical point of view communication paradigms such as software defined radio and cognitive radio have kept many researchers interested in the field of AMC [2] and researchers are proposing new techniques to increase the overall accuracy or reduce complexity.

Gabor filter has been widely adopted in various fields such as image and texture analysis, owing to their excellent feature extraction properties [3]–[5] but the classification performance is comparatively lower. Therefore, the Gabor filtering technique is used here to extract distinct features which are used to classify the modulation formats. These extracted features are then further optimized with a cuckoo search algorithm (CSA), a new meta-heuristics computational technique motivated by parasitism behavior of some species known cuckoo [6]. At the end for decision making on the basis of these optimized features, extreme learning machine (ELM) is utilized. ELM is a new learning algorithm for the single hidden layer feed-forward neural networks. Its learning process is very fast because it is built on empirical risk minimization theory [7].

The main contribution of our research work is the novel approach where instead of using Gabor features directly, we have optimized the features using a meta-heuristic algorithm before classification by a fast-convergent ELM classifier. CSA optimization enhances the accuracy of the algorithm, while ELM increases the efficiency of the classifier and it overcomes the slow training speed and overfitting problems. Taken together, the synergy of the proposed hybrid ELM-CSA methodology resulted in better classification accuracy at comparatively lower SNRs and lower sample size, which is desirable for any communication system. The proposed classifier is tested for both AWGN and Rayleigh fading channels.

The associate editor coordinating the review of this manuscript and approving it for publication was Minho Jo.

This paper is organized in the following manner. A quick review of related work is presented in section II. System model and the main steps of the proposed ELM-CSA approach is detailed in section III. This section describes the extraction of Gabor features, Optimization of these feature through CSA and Classification of modulation schemes through ELM. Training and testing of the algorithm are also discussed in this section. Simulation result at both Fading and non-Fading channel, the accuracy of SCA and probability of correct classification of the proposed algorithm are deliberated in section IV. The paper is then concluded in section V, by presenting a comparison of ELM-CSA method with existing state-of-the-art techniques and future work.

II. RELATED WORK

More than a few research groups have established various modulation recognition methods in the recent past. The paper [8] proposes a method of modulation classification algorithm by considering frequency selective channels and using single and multiple antenna systems. In this paper, they have developed a hypothesis test to detect the correlation-made peaks. The author proposes the direct practice of a similarity measure built on information theory over AWGN channel for the automatic classification of digital modulations schemes, known as the correntropy coefficient in [9]. In [10], the author investigates the technique of DNN by selecting the distinct statistical features over fading and non-fading channel. The method proposed in [11] used a likelihood-based (LB) classifier for SISO system this method is highly computational complex. The author specified an idea of non-data aided channel estimation in [12] for blind modulation classification (MC) in multiple-input-multiple-output (MIMO) fading channels by using maximum likelihood classifier. In [13] they have proposed a method of MC for multiple overlapped sources. Through blind channel estimation overlapped sources are separated then maximum-likelihood-based multi cumulant classification (MLMC) applied to each source. The idea presented in [14] deliberate likelihood-based (LB) statistical tests for AMC in cognitive radio (CR). The author also presented an idea to identify the transmitted replica on the receiving side by using a look-up table (LUT). The method proposed in [15] translates the raw modulated signals into images that have a grid-like topology and feed them to CNN for classification. In [16] authors developed a technique for the examination and classification of the low probability of intercept (LPI) radar waveforms. These signals are classified based on the nature of the pulse compression waveform. [17] proposes an optimized distribution sampling test (ODST) classifier for classification of M-QAM signals. Genetic Algorithm (GA) is used here for optimization of distance matrices. In [18] a method of feature-based AMC proposed for a MIMO system. For separation of signals independent component analysis (ICA) is applied. higher-order cumulants are considered featured and Quasi-Newton method is used for classification purposes.

Authors in [19], not only classified digitally modulated signals in terms of modulation formats and bit-rates but also exploited SNR sensitive features for non-data-aided (NDA) SNR estimation using asynchronous delay-tap plots (ADTPs). The reported classification accuracy of the proposed technique is 99.12% and mean estimation error of 0.88dB. In an attempt to reduce the computational complexity of maximum-likelihood classifier that requires prior knowledge of SNR, a novel idea of first estimating the SNR before classification combined with simplified minimum distance is presented in [20]. Further, a closed form algorithm based on blind source separation is adapted to rectify the carrier phase offset prior to classification procedure. A deep learning approach for modulation classification in AWGN and flat-fading channels is presented. Sparse auto-encoders and supervised SoftMax classifier are employed to achieve good classification accuracy at lower SNRs. For analog modulation classification, authors in [21] have used adaptive wavelet entropy for extracting features and multi-layer perceptron neural network for classification, thus achieving 98.34% classification accuracy. The author proposed AMC framework based on dictionary learning in [22], where first a dictionary is formed using signals with well-known modulation schemes and by using sparse representation on the dictionary, modulation formats of unknown signals are determined then. A genetic programming-based method of modulation classification for overlapped sources (GPOS) is mentioned in [23]. In [24] researches mentioned a technique by utilizing deep convolutional neural networks to classify multicarrier waveforms. FBMC-OQAM, UFMC, and OFDM-QAM are the formats considered in this paper. In [25] convolutional neural network (CNN) based system is built to recognize the cognitive radio waveforms under high power background noise. Modulation classification system under varying noise conditions is proposed in [26] that ensure robustness to SNR variations.

III. SYSTEM MODEL AND PROPOSED CLASSIFIER STRUCTURE

The system model is presented in figure 1. The AMC module consists of three stages: (i) Feature Extractor, (ii) Heuristic Optimizer and (iii) ELM Classifier.

A random signal x(n) is generated on the transmission side; after modulation, it is passed through some pre-defined channel (AWGN/ Rayleigh) with some pre-specified SNR value. The signal sensed on the receiving side represented as:

$$r(n) = x(n) + g(n) \tag{1}$$

where r(n) is the received signal, g (n) is additive white Gaussian noise having zero mean and variance σ_g^2, and x(n) is given by:

$$x\,(n) = \alpha e^{i(w_0 nT+\theta_n)} \sum\nolimits_{j=-\infty}^{j=\infty} x\,(l)\,\rho(nT + j\tau + \varepsilon_T) \tag{2}$$

where $x\,(l)$ is the input sequence, α is signal amplitude, w_0 is angular frequency offset constant, $\rho(..)$ is channel effect, τ is

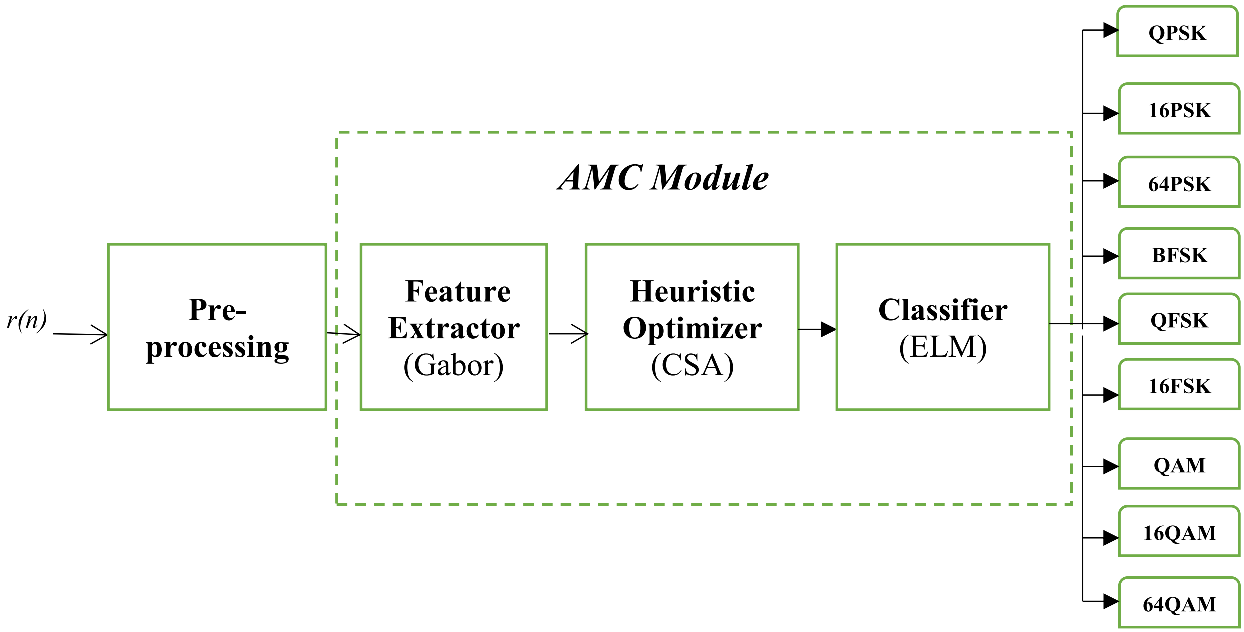

FIGURE 1. Proposed system model.

symbol spacing, ε_T is timing jitter and θ_n is phase jitter which varies from symbol to symbol [27]. The transmitter encodes sequences of randomly generated bits into continuous signal patterns by selecting the appropriate symbols. The received signal arrives in distorted forms suffering from undesirable frequency, timing and phase offsets, that are induced by channel effects and Intersymbol interference (ISI). As described in the system model (figure 1), the received signal is first given to the pre-processing block which is actually used to eradicate all the undesirable effects and the cleaned signal is then fed to AMC module.

Next, we describe the working of the three sub-modules of AMC, i.e., Feature Extractor, Heuristic Optimizer, and ELM classifier.

A. FEATURE EXTRACTION

The first stage in the AMC module is feature extraction where Gabor filter is used to extract the different features for classification of considered digitally modulated schemes. Input that is applied to the Gabor filter is converted into parallel through a serial to parallel converter [27]. These individual input sequences are then applied to the Gabor nodes. Gabor atom computed in [28] is as follows:

$$h_i\left(n_k\right) = \frac{1}{\sqrt{\sigma}_i} g\left(\frac{n_k - C_i}{\sigma_i}\right) e^{jw_i n_k} \quad 1 \leq \mathrm{k} \leq \mathrm{M} \tag{3}$$

where c represent a shift, σ is scale and f is the modulation parameters and $g\left(a\right)$ is defined as:

$$g\left(a\right) = 2^{1/4} e^{-\pi a^2} \tag{4}$$

Thus, the output of the Gabor atom is defined as the inner product of the input sequence and the respective Gabor atom,

$$O_{ki} = \left| < h_i\left(n_k\right), r_{ki} > \right| \tag{5}$$

The output of the Gabor filter is defined as:

$$y_k\left(n\right) = \sum\nolimits_{i=1}^{M} O_{ki}\, w_{ki} \tag{6}$$

where w_{ki} is the weight through which the Gabor node connected to the output node? Our desire is to minimize the square of the error function that is equal to the difference between the desired output d(n) and estimated output y(n), given as:

$$J_k\left(n\right) = \left[d_k\left(n\right) - y_k\left(n\right)\right]^2 \tag{7}$$

TABLE 1. Extraction parameters for Gabor features.

Gabor Features Ranges		
Parameter	**Lower Bound Value**	**Upper bound value**
Scale $\boldsymbol{\sigma}$	1	20
Shift $\boldsymbol{c}$	4	6
Modulation $\boldsymbol{f}$	- π	3π
Weight $\boldsymbol{w}$	0	1

B. HEURISTIC OPTIMIZER

The Gabor features extracted in the previous step is further optimized through CSA.

Cuckoo Search is a relatively new meta-heuristics technique driven by parasitism behavior of some bird species known cuckoo. These birds search nest of other host bird and place their egg in it, nests are selected through some predefined norms that increase the hatching probability of their eggs. If the host bird is not able to recognize the cuckoo eggs, it identifies the egg as its own. Whenever host birds realize that these are the cuckoo eggs, then they will either dispose of these alien eggs or destroy the nest and make a fresh nest at another place. This behavior is utilized to develop the heuristic computational technique known as Cuckoo Search Algorithm, as given in [5]. It consists of the following steps:

1) PARAMETER INITIALIZATION

The parameters are initialized in the first step, i.e., number of nests (n), discovering probability (p_a), the step size parameter (α), and the maximum number of generations as termination criteria.

2) GENERATE INITIAL NESTS

This is the 2nd step where initially Nests are created randomly, and the eggs of host bird are placed in it. Nests are created to find out and update the four unknowns of Gabor filter (σ, c, f, w). Each of them created as:

$$nest_{i,j}^{(0)} = Round\left(l_{j.min} + rand\left(l_{j.max} - l_{j.min}\right)\right) \quad (8)$$

where $nest_{i,j}^{(0)}$ represent the entry of i^{th} row and j^{th} column of respective Nest, $l_{j.min}$ and $l_{j.max}$ are the minimum and maximum limits.

3) GENERATE NEW CUCKOOS BY LÉVY FLIGHTS

All the rows of each Nest represent a solution, separate the row from each Nest that provided the best solution and updates the remaining rows through Lévy Flights [29].

$$nest_i^{(t+1)} = nest_i^{(t)} + \alpha.\mathtt{A}.\left(nest_i^{(t)} - nest_{best}^{(t)}\right).r \quad (9)$$

where $nest_i^{(t+1)}$ represent the i^{th} row of new updated Nest, α is the step size parameter which depends on the nature of the problem determine how far a step of the walker can go, r is a random number from a standard normal distribution, $nest_{best}^{(t)}$ is the best solution so far from the previous Nest and $\mathtt{A}$ is a random walk founded from the Lévy flights.

The step span $\mathtt{A}$ as mentioned in [30] can be calculated as:

$$\mathtt{A} = \frac{u}{|v|^{1/\beta}} = \mathrm{U} \quad (10)$$

where the parameter β can be selected between the interval [1], [2] for this specific problem it is supposed 1.5; u and v from the normal distribution with zero mean and variance σ_u and σ_v respectively can be calculated as:

$$u \sim N(0, \sigma_u^2), \quad v \sim N(0, \sigma_u^2) \quad (11)$$

$$\sigma_u = \left\{\frac{T(1+\beta).\sin(\pi\frac{\beta}{2})}{T\left[\frac{(1+\beta)}{2}\right].\beta.2^{(\beta-1)/2}}\right\}^{1/\beta}, \quad \sigma_V = 1 \quad (12)$$

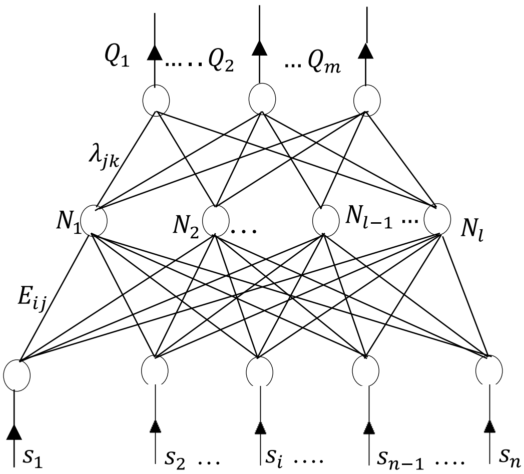

FIGURE 2. ELM classifier structure.

a: GREEDY SELECTION

After the updating of eggs through levy flight some parameters get out of the predefined limits of (σ, c, f, w), it is required to take it back in the required limits. If any entry (egg) is beyond the limit ($nest_{i,j}^{(0)} > l_{j.max} \,||nest_{i,j}^{(0)} < l_{j.min}$) then the following step is needed to use for each entry.

$$nest_{i,j}^{(t+1)} = \frac{(nest_{i,j}^{(t+1)})}{\max\left(nest_{i,j}^{(t+1)}\right) + l_{j.min}} \quad (13)$$

Now again find the best solution of the updated Nest and sort out the Nest. Greedy selection is the process of choosing the best candidate from both (old and updated) Nests, and form a new Nest that contains only Best fit candidate.

b: ALIEN EGGS DISCOVERY

The following discovering probability matrix is considered to expose the alien eggs in the Nest for each solution:

$$P_{ij} = \begin{cases} 1, & rand < P_a \\ 0, & rand \geq P_a \end{cases} \quad (14)$$

where *rand* generates a random number in the interval [0, 1] and P_a is the probability of discovering alien eggs that is predefined.

The eggs in $\mathrm{Nest}_{\mathrm{best\_fit}}$ will be replaced with the new generated one through the random walk that is as follows:

$$nest_{best\_fit}^{(t+1)} = nest_{best\_fit}^{(t)} + Z.P \quad (15)$$

P is the probability matrix given in equation (14) and Z is the step size given as:

$$Z = [rand.(nests\,(randperm1\,(n)\,,:) - nests\,(randperm2\,(n)\,,:))] \quad (16)$$

where *randperm* is the random permutation function applied to the Nests matrix. After this updating function, it is required to repeat the step mentioned in (13) to bring the variable back in limits.

Optimization of Gabor features through CSA

1. **Step I: Parameter Initialization**
   - number of nests (n), discovering probability (p_a), the step size parameter (α)
   - termination criteria.
2. **Step II: *Generate Initial Nests***
   - Set the values of (σ, c, f, w) according to table

for all K
   - Generate Nest randomly according to (7)
   - Calculate $y_k(n)$ through (6) hence finding fitness by (7)

end
   - Rank the solution and finding the best one
3. **Step III: *Generate New Cuckoos by Lévy Flights***
   - *Calculate A, u, v, σ_u, σ_V by using (9),(10) and (11)*
   - Generate nest $nest_i^{(t+1)}$ by using (8)
4. **Step IV: Greedy Selection**

   If
   $(nest_{i,j}^{(0)} > l_{j.max} \,\|\, nest_{i,j}^{(0)} < l_{j.min})$
   - Update values through (13)

   else
   - Go to step V
5. **Step V: *Alien Eggs Discovery***
   - Discovering probability P_{ij}
   - Update $\text{Nest}_{\text{best\_fit}}$ through (15)
   - Repeat step IV
   - Rank the nest by finding fitness & choose current best
   - Keep the best one say, $nest_{best\_fit}^{(t+1)}$
6. **Step VI: *Termination criterion***
   - Repeat III till the completion of a total number of rounds T

FIGURE 3. Gabor feature optimization using CSA.

c: TERMINATION CRITERION

The Step from Generation through Lévy flight to onward are performed alternatively until the termination criteria are satisfied. To update the Gabor features (unknown), each nest related to each unknown and each egg of the respective nest represent a candidate solution. It is required to update each Nest separately, but all the unknown is collectively participating in choosing the best solution. Using the above-stated rules, the elementary steps of the CSA optimization are summarized in figure 3.

IV. ELM CLASSIFIER

Typical feedforward neural network structure elaborated in [6] is shown in Figure 2, which is composed of three layers: input layer S, a single hidden layer N, and output layer Q. This is a fully connected structure between the input to hidden and between hidden to the output layer. The *n* neurons of the input layer corresponding to the input vectors of length n. There are *l* hidden layer neurons and m output layer neurons, corresponding to the *m* output variables [31]. This variable *l*, m, and n are independent of the same previous variables used in this paper. The connection weights between the input layer and the hidden layer are denoted by E.

ELM based modulation Classification

1. **Step I: Training of ELM**
   - **a. Parameter Initialization**
     - Training set = {(Si , Qi), i = 1, · · · ,n }
     - Select Activation function g, and the number of hidden nodes N
     - Generate weight matrix **E** and hidden Node bias **b** randomly
   - **b. The process at Hidden Layer**
     - Calculate the hidden layer output matrix **F** through (23)
     - Hence Calculate the hidden layer to output weight,

       $\boldsymbol{\lambda} = \mathbf{F}^{\dagger}\mathbf{T}'$
2. **Step II: Testing of ELM**
   - **a. Get Values from CSA**
     - Get the best nest from CSA $nest_{best\_fit}^{(t+1)}$
     - Store these values in input Vector S
   - **b. Process at output layer**
     - Calculate the output matrix Q through $\boldsymbol{F\lambda} = \boldsymbol{Q'}$
     - Display the result

FIGURE 4. ELM based modulation classification.

$$E = \begin{bmatrix} E_{11} & E_{12} & \dots & E_{1n} \\ E_{21} & E_{22} & \dots & E_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ E_{l1} & E_{l2} & \dots & E_{ln} \end{bmatrix}_{lXn} \tag{17}$$

where, E_{ij} is the weight between i^{th} neurons of the input layer and j^{th} neurons of the hidden layer.

$\boldsymbol{\lambda}$ is the weight metrics that represent the weight of the respective the hidden layer to the output layer, given as:

$$\boldsymbol{\lambda} = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1m} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_{l1} & \lambda_{l2} & \dots & \lambda_{lm} \end{bmatrix}_{lXm} \tag{18}$$

where λ_{jk} is the connection weights between *k* neurons of the output layer and *j* neurons of the hidden layer. The bias for the hidden neurons *b* is:

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_l \end{bmatrix}_{lx1} \tag{19}$$

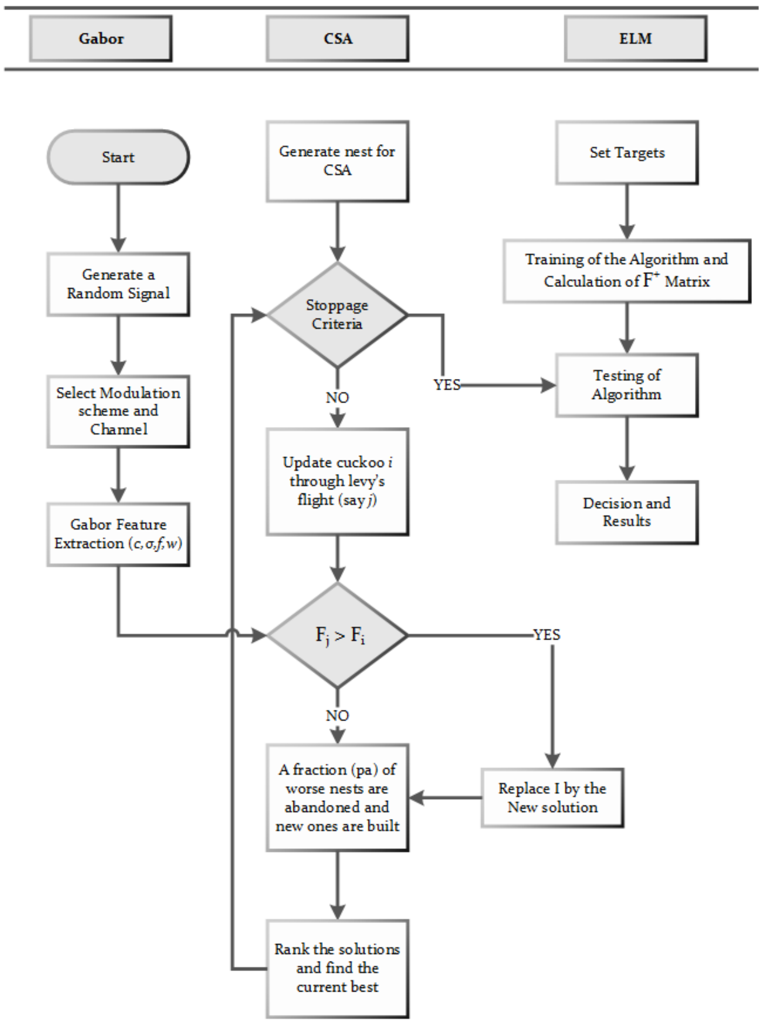

FIGURE 5. **Flow diagram of proposed algorithm showing working of three sub-modules.**

S, the input matrix training set with n samples, and $\boldsymbol{Q}$ the output matrix training set with m samples respectively are given by:

$$S = \begin{bmatrix} s_{11} & s_{12} & \dots & s_{1Q} \\ s_{21} & s_{22} & \dots & s_{2Q} \\ \vdots & \vdots & \vdots & \vdots \\ s_{nl} & s_{n2} & \dots & s_{nQ} \end{bmatrix}_{nXl} \tag{20}$$

The hidden layer neuron activation function is $g(s)$, from Figure 2 we can get the network output $\boldsymbol{Q}$ [32].

$$\mathbf{Q} = [\boldsymbol{Q}_1, \quad \boldsymbol{Q}_2, \quad \dots, \quad \boldsymbol{Q}_l\ ]_{mXl} \tag{21}$$

$$\boldsymbol{Q}_j = \begin{bmatrix} Q_{1j} \\ Q_{2j} \\ \vdots \\ Q_{mj} \end{bmatrix}_{mx1} = \begin{bmatrix} \sum_{i=1}^{l} \lambda_{i1} g(E_i s_i + b_i) \\ \sum_{i=1}^{l} \lambda_{i2} g(E_i s_i + b_i) \\ \vdots \\ \sum_{i=1}^{l} \lambda_{im} g(E_i s_i + b_i) \end{bmatrix}_{mx1} \tag{22}$$

where j $= (1, 2, \dots, l)$, $E_i = [E_{i1}, E_{i2}, \dots, E_{in}]$, $s_j = [s_{1j}, s_{2j}, \dots, s_{nj}]^T$

Equation (22) can be expressed as $\boldsymbol{F\lambda} = \boldsymbol{Q}$, where $\boldsymbol{Q}$ is the transpose of matrix $\boldsymbol{Q}$, $\mathbf{F}$ is the output matrix of the neural network hidden layer [4].

$$\boldsymbol{F}(E_1, E_2, \dots, E_l, b_1, b_2, \dots, b_l, s_1, s_2, \dots, s_n) = \begin{bmatrix} g(E_1 \bullet s_1 + b_1) & g(E_2 \bullet s_1 + b_2) & g(E_l \bullet s_1 + b_l) \\ g(E_1 \bullet s_2 + b_1) & g(E_2 \bullet s_2 + b_2) & g(E_l \bullet s_2 + b_l) \\ & \vdots & \\ g(E_1 \bullet s_n + b_1) & g(E_2 \bullet s_n + b_2) & g(E_l \bullet s_n + b_l) \end{bmatrix}_{nxl} \tag{23}$$

The connection weights $\boldsymbol{\lambda}$, between the hidden layer and the output layer, can be obtained by using the least squares method to solve the following equation.

$$\min \| \boldsymbol{F\lambda} - \boldsymbol{O'} \| \tag{24}$$

The solution is $\boldsymbol{\lambda} = \mathbf{F}\dagger\mathbf{T}'$, where $\mathbf{F}\dagger$is the Moore Penrose generalized inverse of the hidden layer output matrix. Steps at the 3rd stage of the system model are summarized in figure 4.

Summary of the CSA-ELM Algorithm: The flow diagram in figure 5 depicts the step-wise methodology of

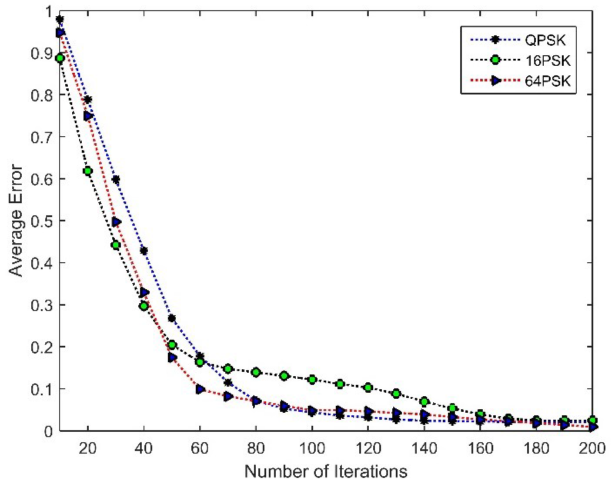

FIGURE 6. Avg. error vs. Iterations plot for variants of PSK using CSA on AWGN channel.

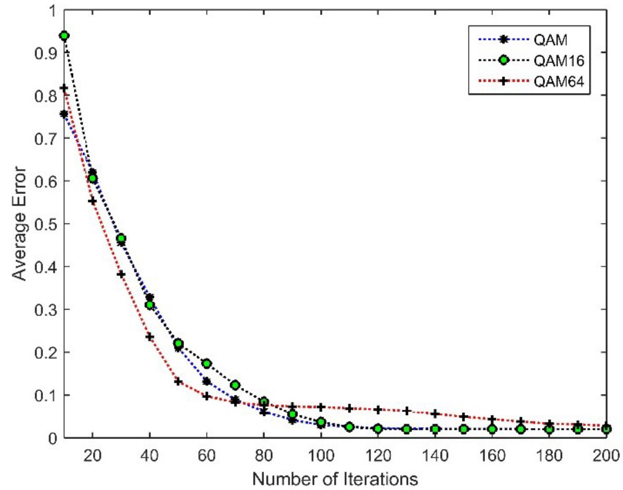

FIGURE 8. Avg. error vs. Iterations plot for variants of QAM using CSA on AWGN channel.

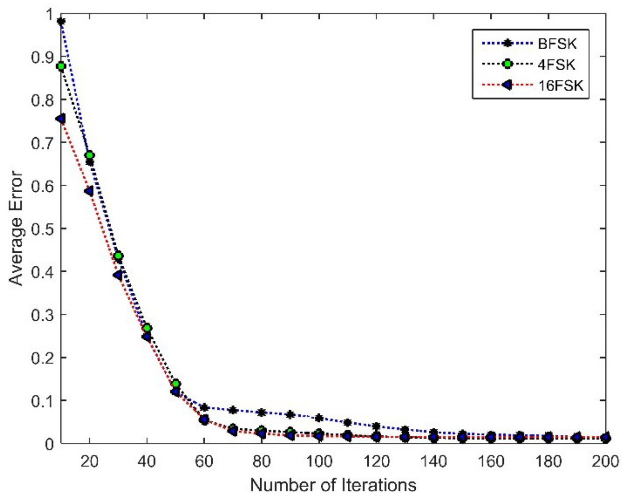

FIGURE 7. Avg. error vs. Iterations plot for variants of FSK using CSA on AWGN channel.

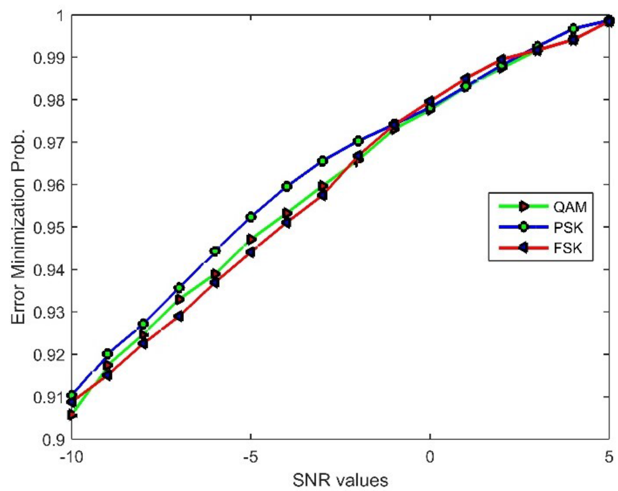

FIGURE 9. Comparison of error minimization probability for FSK, PSK, and QAM using CSA on AWGN channel.

the algorithm. The working of three core modules, i.e., Gabor, CSA, and ELM of the proposed system can be seen in parallel to each other. Gabor feature extraction module extracts Gabor features (c, σ, f, w) from randomly generated signal passed through either of the two channels as depicted in equation (5). The extracted Gabor features are distinct but to achieve better classification accuracy, they are further optimized using CSA using the fitness function shown in equation (7). The best solution to having max fitness is then fed to ELM classifier. The ELM classifier (already trained according to reference values) then makes a decision using equation (24) about the modulation classification.

V. SIMULATIONS, RESULTS AND DISCUSSION

In this section, we analyze the performance of our proposed CSA optimizer and ELM classifier by conducting a range of experiments using exhaustive Monte-Carlo simulations for total nine different variants of PSK, FSK and QAM considering different sample sizes (512 and 1024) and SNRs over AWGN and Rayleigh fading channels. The simulations are performed on MATLAB R2015 running Microsoft Windows 10 on HP Core- i3 workstation with 4GB RAM.

Before we proceed to results, lets us first define the following performance metrics that are later used in the detailed analysis. *Error Minimization Probability (EMP)* is used to compare the effectiveness of CSA for different modulation scheme at different SNRs. It depicts how quickly is the probability of error between the desired and estimated features of CSA at a specific SNR is minimized (Lower the probability of error, higher will be the EMP). Average error or Average EMP is used for comparing the performance of CSA of variants of same modulation schemes (e.g., QAM, 16-QAM, and 64-QAM). To evaluate the efficiency of CSA for all the modulation schemes under consideration for a different set of SNRs, we have used Error Optimization Accuracy.

TABLE 2. Error optimization accuracy of CSA over AWGN channel.

Modulation Schemes	Error Optimization Accuracy of CSA				
	SNR	-20 dB	-10 dB	-5 dB	0 dB
QPSK		89.80%	92.90%	97.50%	99.50%
16PSK		89.15%	92.30%	94.15%	98.70%
64PSK		87.61%	91.26%	95.53%	98.85%
BFSK		88.66%	92.80%	97.94%	99.70%
QFSK		87.35%	91.43%	94.80%	98.90%
16FSK		86.15%	90.09%	93.50%	98.59%
QAM		87.50%	89.60%	95.62%	98.14%
16QAM		88.43%	90.40%	95.90%	99.05%
64QAM		85.25%	88.52%	95.40%	98.65%

TABLE 3. Percentage classification accuracy on AWGN channel.

	Modulation Schemes	Percentage Classification Accuracy								
		SNR 0 dB								
		QPSK	16PSK	64PSK	BFSK	QFSK	16FSK	QAM	16QAM	64QAM
512 Samples	QPSK	99.8%								
	16PSK		99.7%							
	64PSK			99.9%						
	BFSK				99.7%					
	QFSK					99.9%				
	16FSK						99.8%			
	QAM							99.8%		
	16QAM								99.8 %	
	64QAM									99.7 %
1024 Samples	QPSK	100 %								
	16PSK		100 %							
	64PSK			100 %						
	BFSK				99.8%					
	QFSK					100 %				
	16FSK						100 %			
	QAM							100 %		
	16QAM								99.9 %	
	64QAM									100 %

Percentage Classification Accuracy (PCA) is used to validate the performance of ELM classifier.

A. PERFORMANCE OVER NON-FADING CHANNEL

In figures [6, 7, 8], the average error is plotted against a number of iterations for different variants of PSK, FSK, and QAM, respectively at 0dB SNR taking 512 sample size and a maximum of 200 iterations.

It can be observed that for PSK and QAM specifically, the convergence is better for modulations with an increasing number of constellations/orders. For FSK, there is a sharp drop in avg error for 4-FSK at approximately 80 iterations but around 140 iterations the BFSK takes over and in the longer run feature optimization is slightly better for 16-FSK. As a generic trend, avg error is reduced with an increasing number of iterations and almost at ∼150 number of iterations, all the schemes become near to equal in term of avg error. Figure 9 shows the plot of EMP against SNRs [−10, −5,0,5] dB. Overall, it can be seen that at low SNRs [−10 to 0dB], EMP is better for PSK. From 0 to 2.5dB SNR FSK, EMP is slightly better and the proposed solution gives almost max EMP at 5 dB.

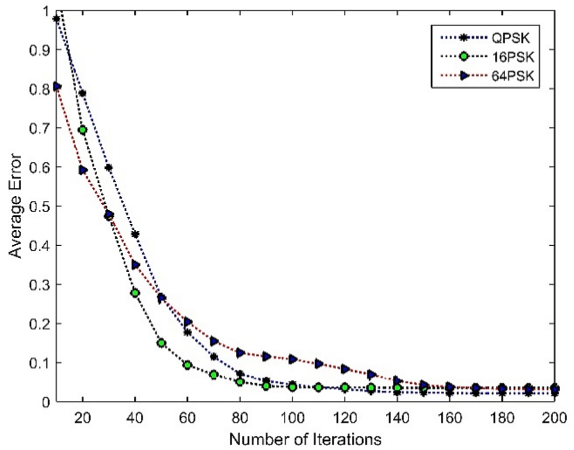

FIGURE 10. Avg. error vs. iteration plot for PSK using CSA on fading channel.

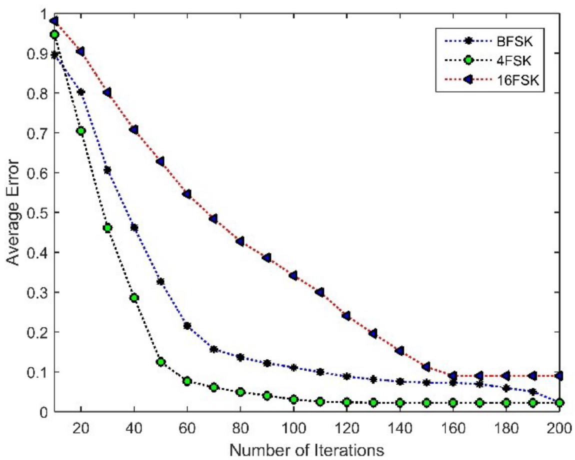

FIGURE 11. Avg. error vs. iteration plot for FSK using CSA on fading channel.

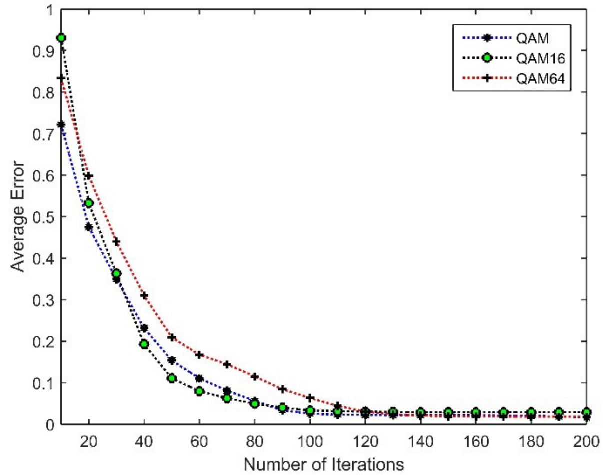

FIGURE 12. Avg. error vs. iteration plot for QAM using CSA on fading channel.

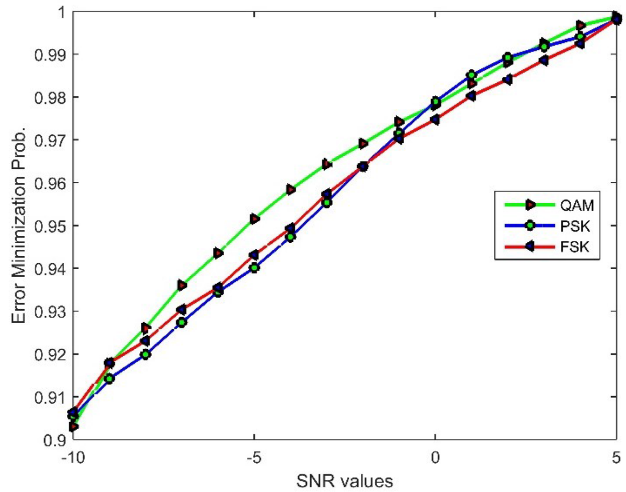

FIGURE 13. Comparison of error minimization probability for FSK, PSK, and QAM using CSA on Rayleigh channel.

Table 2 shows the error optimization accuracy of CSA over the AWGN channel. The results of the different modulation schemes used have been mentioned for different SNR values for 512 samples. From the table, it is quite evident that the methodology applied has minimized the error to a distinguishable extent. In the PSK modulation scheme, our CSA performs better for QPSK than the 16PSK and 64PSK. At lower SNRs [−20, −10 dB] 16PSK optimization accuracy is better but at −5dB and 0dB, 64PSK is better as compared with 16PSK. In FSK, the accuracy level for BFSK is better than the QFSK and 16FSK but only by a slight margin. Likewise, in the QAM modulation scheme, CSA shows better optimization for 16QAM compared to QAM and 64-QAM, owing to a regularized change of percentage. Overall, PSK shows distinguishable results than PSK and FSK.

Table 3 manifests the overall classification accuracy (PCA) of our proposed solution for all the considered modulation schemes with different samples sizes (512, 1024) at 0 dB SNR for the AWGN channel. Here we have considered the 1000 trails of ELM, and calculated results have been shown in the respective tables. Almost all the modulations schemes are classified with an accuracy of ∼99.7% at 512 samples, which becomes ∼100% for 1024 sample size.

B. PERFORMANCE OVER FADING CHANNELS

For fading channel, we have considered the Rayleigh fading as a worst-case scenario at 1000 symbol/sec with sampling period 1ms and maximum Doppler shift $\leq$ $[1/(10*T_s)]$, where T_s is the sampling period. Figures [10-12] show the optimized features result of PSK, FSK and QAM through CSA of various order at 0 dB SNR value over the considered Rayleigh fading channel.

Among PSK variants (Figure 10), 16-PSK converges quite earlier (∼80-85 iteration mark) towards optimum value, but around ∼140 iterations, avg error for the other two variants (QPSK and 16-PSK) becomes almost same to 64-PSK.

In figure 11, there is a close contest between BFSK and 4-FSK as compared to 16FSK optimization through our

TABLE 4. Error optimization accuracy of CSA over fading channel.

Modulation Schemes	Error Optimization Accuracy of CSA			
SNR	-20 dB	-10 dB	-5 dB	0 dB
QPSK	91.39%	93.54%	95.57%	98.90%
16PSK	89.23%	91.27%	94.46%	98.12%
64PSK	88.43%	90.52%	93.74%	98.09%
BFSK	88.42%	91.45%	95.13%	97.95%
4FSK	88.63%	92.25%	95.56%	98.62%
16FSK	82.85%	87.31%	92.67%	96.34%
QAM	87.82%	91.67%	97.46%	99.49%
16QAM	86.21%	89.78%	93.35%	98.20%
64QAM	87.24%	90.08%	93.66%	98.67%

TABLE 5. Percentage classification accuracy on fading channel.

	Modulation Schemes	Percentage Classification Accuracy								
	SNR	0 dB								
		QPSK	16PSK	64PSK	BFSK	QFSK	16FSK	QAM	16QAM	64QAM
512 Samples	QPSK	99.6%								
	16PSK		99.9%							
	64PSK			100%						
	BFSK				99.8%					
	QFSK					98.6%				
	16FSK						99.3%			
	QAM							99.9%		
	16QAM								99.5 %	
	64QAM									99.8 %
1024 Samples	QPSK	100 %								
	16PSK		100 %							
	64PSK			100 %						
	BFSK				99.9%					
	QFSK					100 %				
	16FSK						99.9%			
	QAM							100 %		
	16QAM								100 %	
	64QAM									100 %

proposed CSA based solution. BFSK converged very slow. Around ∼200 iterations, avg error in desired and estimated features for BFSK and 4FSK modulation schemes is almost the same.

For QAM modulation in figure 12. A pattern similar to PSK can be observed that the avg error for higher order variant 16-QAM falls quickly towards optimum at ∼80 iteration mark, but all performance of the CSA optimizer becomes same for all the QAM variant around ∼120 iteration mark. However, the avg error is much lower for QAM and 64-QAM at iterations above 120, and the difference is not quite distinguishable. These results are obtained on the Rayleigh channel by considering the 0 dB SNR values.

An average error minimization probability for all the modulation variants is shown in Figure-13 Starting with −10dB SNR value. It can be seen that with the increase in SNR values graph converges toward the maxima and at 5 dB SNR probability reached almost it's the maximum value.

Table 4 exhibits the error optimization accuracy of CSA over the Fading channel for the mentioned SNR values using 512 samples. It is evident that in the PSK modulation scheme, QPSK has better error minimization percentage than 16PSK and 64PSK. The values, however, only differ by a minute ratio.

In FSK modulation, 4FSK shows more promising results as compared to BFSK and 16FSK at lower SNRs. The values

TABLE 6. Performance evaluation table for the comparison with other techniques.

Modulation Schemes	Accuracy at Respective SNR values					
	[36] SNR = 0 dB	[33] SNR=0,10 dB	[34] SNR=5dB.	[35] SNR=0dB	[28] SNR=0dB	Proposed algo. SNR = 0dB
	KNN Classifier	Polynomial classifier	VLR Classifier	HMM Classifier	Gabor Filter Network	ELM Classifier
QPSK	99.97%	70.7%	95%	--	63 %	100%
16PSK	-- --	--	19%	--	--	100%
64PSK	--	--	28%	--	--	100%
BFSK	--	--	--	--	``	99.8%
QFSK	--	--	--	--	--	100%
16FSK	--	--	--	--	58 %	100%
QAM	90.64%	--	--	72.35%	--	100%
16QAM	99.89%	95.1%	--	71.94%	--	99.9%
64QAM	99%	95.1%	--	69.96%	--	100%

for BFSK do not change drastically like in the case of 4FSK and 16FSK. In the case of QAM modulation, 16QAM lags behind in the error minimization as compared to QAM and 64QAM at lower SNRs. Also, between QAM and 64QAM, CSA shows better results for the later at lower SNRs, but as the SNRs is increased, the error optimization accuracy for QAM is visibly better.

Table 5 displays the percentage classification accuracy (PCA) of our proposed CSA-ELM classifier for different variants of PSK, FSK and QAM considering samples sizes (512, 1024) at 0 dB SNR for Rayleigh channel. Here we have considered the 1000 trails of ELM, and calculated results have been shown in the respective tables. Almost all the modulations schemes are classified with an accuracy of ∼99% at 512 samples, which becomes ∼100% for 1024 sample size.

Comparing Tables 3 and 5, that shows the PCA for our proposed solutions for considered modulations schemes over both fading and non-fading channels, the classification accuracy is almost 100% at 0dB SNR and 1024 sample size. This validates the effectiveness of using CSA and ELM for our proposed solution to this problem consideration at lower SNRs and lower sample size.

C. COMPARISON WITH EXISTING TECHNIQUES

Table 6 compares the performance of the proposed algorithm with existing literature having different classifiers for 1024 samples. In [28] the authors have used Gabor filter network for the classification with common modulation schemes as QPSK and 16FSK at 0 dB SNR having a meager accuracy of 63% and 58% only. Cepstral features with VLR classifier have been used by authors in [34] for modulation classification. It can be seen that even at 5dB SNR the classification accuracy is no way near our accuracy of 100% at 0dB for common modulations of QPSK, 16PSK, and 64PSK. Authors in [35], have used a novel approach for Automatic Modulation Classification via Hidden Markov Models and Gabor Features for the common modulation schemes of QAM, 16QAM and 64QAM with an accuracy of 72.35%, 71.94% and 69.96% at 0 dB. Comparing the results of above-mentioned techniques with our ELM-CSA approach, the results clearly manifests the supremacy of our proposed algorithm.

To further highlight the effectiveness of our proposed approach, we have compared the results with two more articles [33] and [36] in which the classification is also done using cummulants. It can be observed that our ELM-CSA based classification approach outperforms the other two classifiers.

VI. CONCLUSION AND FUTURE WORK

In this paper, we have presented a novel approach of using Extreme Learning Machine for classification of modulations schemes by extracting the Gabor Features and Optimizing them through Cuckoo Search Algorithm. Gabor filter provided modulation features in terms of scale, shift and weights that are optimized through CSA by eliminating unwanted channels effects. The proposed solution then exploits the inherent fast learning capability of ELM. Combination of ELM with CSA eventually enhance the classification accuracy. To highlight the effectiveness of our proposed solution, the results are compared with relevant published work.

It is quite evident that the proposed algorithm provides better accuracy for all the considered modulation schemes with a smaller number of samples and at low SNR values. Moreover, the considered modulation set is larger than the one used in these aforementioned research studies, which shows the comprehensiveness of our work.

We have utilized Gabor features in our proposed solution in combination with machine learning and heuristic techniques. Performance of other features such as Cyclostationary and Spectral features can further be explored. Similarly, Deep learning approach instead of ELM can be used with other comparatively new bio-inspired heuristic techniques such as Cat swarm optimization, Bat optimization, firefly optimization algorithms. It will also be interesting to analyze the effect of phase, frequency and time offset on classification accuracy.

REFERENCES

[1] D.-C. Chang and P.-K. Shih, "Cumulants-based modulation classification technique in multipath fading channels," *IET Commun.*, vol. 9, no. 6, pp. 828–835, 2015.

[2] A. Hussain, M. F. Sohail, S. Alam, S. A. Ghauri, and I. M. Qureshi, "Classification of M-QAM and M-PSK signals using genetic programming (GP)," *Neural Comput. Appl.*, pp. 1–9, Mar. 2018.

[3] P. Parida and N. Bhoi, "2-D Gabor filter based transition region extraction and morphological operation for image segmentation," *Comput. Elect. Eng.*, vol. 62, pp. 119–134, Aug. 2017.

[4] N. Nabizadeh and M. Kubat, "Brain tumors detection and segmentation in MR images: Gabor wavelet vs. statistical features," *Comput. Elect. Eng.*, vol. 45, pp. 286–301, 2015.

[5] S. Meshgini, A. Aghagolzadeh, and H. Seyedarabi, "Face recognition using Gabor-based direct linear discriminant analysis and support vector machine," *Comput. Elect. Eng.*, vol. 39, no. 3, pp. 727–745, 2013.

[6] A. H. El-Maleh, S. M. Sait, and A. Bala, "State assignment for area minimization of sequential circuits based on cuckoo search optimization," *Comput. Elect. Eng.*, vol. 44, pp. 13–23, May 2015.

[7] G.-X. Li, "Application of extreme learning machine algorithm in the regression fitting," in *Proc. Int. Conf. Inf. Syst. Artif. Intell. (ISAI)*, Jun. 2016, pp. 419–422.

[8] M. Marey and O. A. Dobre, "Blind modulation classification algorithm for single and multiple-antenna systems over frequency-selective channels," *IEEE Signal Process. Lett.*, vol. 21, no. 9, pp. 1098–1102, Sep. 2014.

[9] A. I. R. Fontes, A. de M. Martins, L. F. Q. Silveira, and J. C. Principe, "Performance evaluation of the correntropy coefficient in automatic modulation classification," *Expert Syst. Appl.*, vol. 42, no. 1, pp. 1–8, Jan. 2015.

[10] B. Kim, J. Kim, H. Chae, D. Yoon, and J. W. Choi, "Deep neural network-based automatic modulation classification technique," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2016, pp. 579–582.

[11] S. Arya, "An efficient likelihood based automatic modulation classification for SISO and MIMO wireless communication systems," Ph.D. dissertation, Nat. Inst. Technol., Rourkela, India, 2017.

[12] Z. Zhu and A. K. Nandi, "Modulation classification in MIMO fading channels via expectation maximization with non-data-aided initialization," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2015, pp. 3014–3018.

[13] S. Huang, Y. Yao, Z. Wei, Z. Feng, and P. Zhang, "Automatic modulation classification of overlapped sources using multiple cumulants," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 6089–6101, Jul. 2016.

[14] A. O. A. Salam, R. E. Sheriff, S. R. Al-Araji, K. Mezher, and Q. Nasir, "A unified practical approach to modulation classification in cognitive radio using likelihood-based techniques," in *Proc. IEEE 28th Can. Conf. Elect. Comput. Eng.*, May 2015, pp. 1024–1029.

[15] S. Peng, H. Jiang, H. Wang, H. Alwageed, and Y.-D. Yao, "Modulation classification using convolutional neural network based deep learning model," in *Proc. 26th Wireless Opt. Commun. Conf. (WOCC)*, Apr. 2017, pp. 1–5.

[16] T. R. Kishore and K. D. Rao, "Automatic intrapulse modulation classification of advanced LPI radar waveforms," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 2, pp. 901–914, Apr. 2017.

[17] A. Ali, F. Yangyu, and S. Liu, "Automatic modulation classification of digital modulation signals with stacked autoencoders," *Digit. Signal Process.*, vol. 71, pp. 108–116, Dec. 2017.

[18] Y. Nie, X. Shen, S. Huang, Y. Zhang, and Z. Feng, "Automatic modulation classification based multiple cumulants and quasi-Newton method for MIMO system," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–5.

[19] U. R. Acharya, H. Fujita, M. Adam, O. S. Lih, V. K. Sudarshan, T. J. Hong, J. E. Koh, Y. Hagiwara, C. K. Chua, C. K. Poo, and T. R. San, "Automated characterization and classification of coronary artery disease and myocardial infarction by decomposition of ECG signals: A comparative study," *Inf. Sci.*, vol. 377, pp. 17–29, Jan. 2017.

[20] J. Zhang, Y. Li, and J. Yin, "Modulation classification method for frequency modulation signals based on the time–frequency distribution and CNN," *IET Radar, Sonar Navigat.*, vol. 12, no. 2, pp. 244–249, Feb. 2017.

[21] D. Avci, "An intelligent system using adaptive wavelet entropy for automatic analog modulation identification," *Digit. Signal Process.*, vol. 20, no. 4, pp. 1196–1206, 2010.

[22] K. Zhang, E. L. Xu, Z. Feng, and P. Zhang, "A dictionary learning based automatic modulation classification method," *IEEE Access*, vol. 6, pp. 5607–5617, 2018.

[23] S. Huang, Y. Jiang, X. Qin, Y. Gao, Z. Feng, and P. Zhang, "Automatic modulation classification of overlapped sources using multi-gene genetic programming with structural risk minimization principle," *IEEE Access.*, vol. 6, pp. 48827–48839, 2018.

[24] S. Duan, K. Chen, X. Yu, and M. Qian, "Automatic multicarrier waveform classification via PCA and convolutional neural networks," *IEEE Access*, vol. 6, pp. 51365–51373, 2018.

[25] M. D. Ming Zhang and L. Guo, "Convolutional neural networks for automatic cognitive radio waveform recognition," *IEEE Access*, vol. 5, pp. 11074–11082, 2017.

[26] Z. Wu, S. Zhou, Z. Yin, B. Ma, and Z. Yang, "Robust automatic modulation classification under varying noise conditions," *IEEE Access*, vol. 5, pp. 19733–19741, 2017.

[27] S. A. Ghauri and I. M. Qureshi, "M-PAM signals classification using modified Gabor filter network," *Math. Problems Eng.*, vol. 2015, pp. 1–10, Apr. 2015.

[28] S. A. Ghauri, I. M. Qureshi, T. A. Cheema, and A. N. Malik, "A novel modulation classification approach using Gabor filter network," *Sci. World J.*, vol. 2014, Jul. 2014, Art. no. 643671.

[29] W. Buaklee and K. Hongesombut, "Optimal DG allocation in a smart distribution grid using cuckoo search algorithm," in *Proc. 10th Int. Conf. Elect. Eng./Electron., Comput., Telecommun. Inf. Technol.*, May 2013, vol. 11, no. 2, pp. 1–6.

[30] X.-S. Yang and S. Deb, "Engineering optimisation by cuckoo search," *Int. J. Math. Model. Numer. Optim.*, vol. 1, no. 4, pp. 330–343, 2010.

[31] G. X. Li, "Application of extreme learning machine algorithm in the regression fitting," in *Proc. Int. Conf. Inf. Syst. Artif. Intell. (ISAI)*, Jun. 2016, pp. 419–422.

[32] Z. B. Duranay and H. Guldemir, "Extreme learning machine based selected harmonic elimination for single phase inverters," *Measurement*, vol. 131, pp. 300–308, 2019.

[33] A. Abdelmutalab, K. Assaleh, and M. El-Tarhuni, "Automatic modulation classification using polynomial classifiers," in *Proc. IEEE 25th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Sep. 2014, pp. 806–810.

[34] M. E.-H. M. Keshk, M. A. El-Naby, R. M. Al-Makhlasawy, H. A. El-Khobby, W. Hamouda, M. M. A. Elnaby, E.-S. M. El-Rabaie, M. I. Dessouky, S. A. Alshebeili, and F. E. Abd El-Samie, "Automatic modulation recognition in wireless multi-carrier wireless systems with cepstral features," *Wireless Pers. Commun.*, vol. 81, no. 3, pp. 1243–1288, 2014.

[35] S. A. Ghauri, I. M. Qureshi, and A. N. Malik, "A novel approach for automatic modulation classification via hidden Markov models and Gabor features," *Wireless Pers. Commun.*, vol. 96, no. 3, pp. 4199–4216, 2017.
[36] A. Hussain, S. A. Ghauri, M. F. Sohail, S. A. Khan, and I. M. Qureshi, "KNN based classification of digital modulated signals," *IIUM Eng. J.*, vol. 17, no. 2, pp. 71–82, 2016.

SYED IHTESHAM HUSSAIN SHAH received the B.S. degree in telecommunication engineering from the National University of Modern Languages (NUML), Islamabad, Pakistan, in 2014, and the M.S. degree in electrical engineering (specialization in signal processing) from International Islamic University, Islamabad, in 2018. He has been a Lab Engineer with the Department of Electrical Engineering, NUML. His research interests include automatic modulation classification, signal processing, and application of heuristic computation techniques in engineering problems.

SHERAZ ALAM received the B.S. degree in electrical engineering from the University of Engineering and Technology, Taxila, Pakistan, in 2002, the M.S. degree in personal, mobile, and satellite communications from the University of Bradford, U.K., in 2005, and the Ph.D. degree in electrical engineering from International Islamic University, Islamabad, Pakistan, in 2018. He has been an Assistant Professor with the Department of Electrical Engineering, National University of Modern Languages, Islamabad. He has more than 15 publications in leading recognized international journals and international and local conference proceedings. His research interests include cognitive radio networks, smart grid communication, automatic modulation classification, and application of heuristic computation techniques in engineering problems. He is a Reviewer in many international journals from the science and engineering domain.

SAJJAD A. GHAURI received the B.Sc. and M.S. degree (specialization in signal processing in communication), in 2007 and 2010, respectively, and the Ph.D. degree in electronic engineering from ISRA University, Pakistan, in 2015, where he has been an Associate Professor, since 2018. Before joining ISRA University, he has been an Assistant Professor with the Department of Electrical Engineering, International Islamic University, Islamabad, Pakistan. He has been conducting research in signal processing, communication, heuristic techniques, and neural network for the past ten years. He is working on modulation classification algorithms, smart grid communication systems, and application of fractional algorithms in massive multiple-input-multiple-output (MIMO) communication systems. He has more than 45 publications in leading recognized international journals and international and local conference proceedings. He is a member of several journal and conference program committees.

ASAD HUSSAIN received the M.Sc. degree in electronics from Quaid-e-Azam University, Islamabad Pakistan, in 2008, and the M.S. degree in electronics engineering from ISRA University, Islamabad, in 2016. He has been a Lecturer with the Department of Engineering, National University of Modern Languages. His research interests include automatic modulation classification, signal processing, and machine learning application of heuristic computation techniques in engineering problems.

FARAZ AHMED ANSARI received the B.Sc. degree (Hons.) in electronic engineering from International Islamic University, Islamabad, in 2008, and the P.G.D. degree in telecom engineering from UET Taxila, in 2015. He is currently pursuing the M.S. degree in electrical engineering from Comsat University, CIIT WAH Cantt, Pakistan. He has been a Lecturer with the Department of Electrical Engineering, National University of Modern Languages, Islamabad, Pakistan, since 2010. His research interests include cognitive radio networks, wireless networks, wireless sensor networks, and the IoT.