



Particle swarm optimized extreme learning machine for feature classification in power quality data mining

S. Vidhya^a and V. Kamaraj^b

^aSri Lakshmi Ammal Engineering College, Chennai, India; ^bSSN College of Engineering, Chennai, India

ABSTRACT

This paper proposes enhanced particle swarm optimization (PSO) with craziness factor based extreme learning machine (ELM) for feature classification of single and combined power quality disturbances. In the proposed method, an S-transform technique is applied for feature extraction. PSO with craziness factor is applied to adjust the input weight and hidden biases of ELM. To test the effectiveness of the proposed approach, eight possible combinations of single and combined power quality disturbances are assumed in the sampled form and the performance of the proposed approach is investigated. In addition white gaussian noise of different signal-to-noise ratio is added to the signals and the performance of the algorithm is analysed. The results indicate that the proposed approach can be effectively applied for classification of power quality disturbances.

ARTICLE HISTORY

Received 3 March 2017
Accepted 9 May 2018

KEYWORDS

Power quality; extreme learning machine; particle swarm optimization; feature classification

1. Introduction

A major challenge in the area of power systems is monitoring and detection of power quality disturbances [1]. The power quality monitoring equipment should be capable of recognizing, capturing and classifying the power quality disturbances [1,2]. In the past few years, researchers have oriented towards the development of feature selection and classification algorithms that can effectively detect power quality disturbances. Techniques like discrete wavelet transform, S-transform and Hilbert transform have been applied for feature extraction [3–9]. Analysis of the literature indicates that, among the available feature extraction techniques, the performance of S-transform is attractive for the signal with noise disturbances. S-transform preserves the phase information during decomposition, as it uses variable window length and engages the Fourier transform kernel [3]. With respect to classification algorithms, researchers have proposed rule based classification, fuzzy logic based classification and neural-based approaches. Of the different classification methods, the neural network based approaches have been widely applied for classification of power quality disturbances [10–16]. The gradient-based learning algorithms, such as back-propagation and its variant Levenberg–Marquardt method, have been applied for training the multilayer feed forward network. Even though the performance of these algorithms is satisfactory, the major problem in these algorithms is that they are slow in learning, get stuck in local minimum and require appropriate choice of stopping criteria, learning rate and learning epochs [17]. In the recent years

there has been a growing research interest towards the analysis of extreme learning machine (ELM), applied for single-hidden-layer feed forward neural networks. ELM, introduced by Huang et al., is a fast learning algorithm where the input weights and hidden biases are randomly generated and the output weights are analytically determined using the least square method. Due to its fast learning and good generalization ability, ELM has been successful in a wide variety of applications [18–20]. However, the random choice of input weights and biases created an uncertainty in regression and classification problems [19]. To alleviate this problem researchers have focused on optimization techniques for choosing the number of neurons in the hidden layer and tuning of input weights and hidden biases [17,21,22]. The authors in [18] have proposed a real-coded genetic algorithm approach to select the optimal number of hidden nodes, input weight and biases. The proposed algorithm B-ELM [23] optimizes the output layer weights by Bayesian linear regression. An improved particle swarm optimization (PSO) algorithm has been proposed to enhance the performance of ELM [24]. From the detailed analysis of the literature [18,19,20,23–25,26], it is evident that optimization techniques have been applied to enhance the performance of ELM. On the subject of optimization algorithms, several variants have been proposed with the objective of evading local minima and improve the search performance. In this perspective, this work focuses on the application of chaotic PSO-based ELM for feature classification. This method incorporates a craziness operator with

chaotic weight updation to maintain diversity in the particles.

The classification accuracy of the proposed method is tested by considering a total of eight power quality disturbances namely, Sag (S1), Harmonic (S2), Interruption (S3), Swell (S4), Swell with Harmonic (S5), Sag with Harmonic (S6), Flicker (S7) and Oscillatory Transient (S8).

The paper traces its flow in eight sections with Section 1 explaining the significance of power quality disturbance classification and the merits and challenges in ELM-based classifier. While Section 2 details the feature extraction using S-transform, Section 3 discusses the basics of ELM classifier. Sections 4 and 5 focus on the principles of PSO and enhanced PSO algorithm, respectively, while Section 6 details the application of chaotic PSO for adjusting the weights and bias of ELM. The results are analysed in Section 7. Finally, the concluding remarks are given in Section 8.

2. Feature extraction using S-transform

The S-transform [27] provides necessary information for analysing and detecting power quality events [11]. This method combines the short-term Fourier transform and the wavelet transform. The derivation of the S-transform is done from the wavelet transform by modifying the phase of the window function or mother wavelet. For a signal $x(t)$, the S-transform is defined as

$$S(\tau, f) = \int_{-\infty}^{\infty} x(t)g_f(\tau - t)e^{-i2\pi ft}, \quad (1)$$

where $g_f(\tau)$ is the gaussian modulation function, defined as follows:

$$g_f(\tau) = \frac{|f|}{\sqrt{2\pi}} e^{-(\tau^2/2\sigma^2)} \quad (2)$$

where $f \in$ denotes the linear frequency and,

$$\sigma = \frac{1}{|f|}. \quad (3)$$

Substituting (2) and (3) in (1) the expression becomes

$$S(\tau, f) = \int_{-\infty}^{\infty} x(t) \frac{|f|}{\sqrt{2\pi}} e^{-\left(\frac{(\tau-t)^2 f^2}{2}\right)} e^{-i2\pi ft}. \quad (4)$$

The discrete version is calculated from the fast Fourier transform. The discrete Fourier transform of the time series $x(t)$ is obtained as follows:

$$H\left[\frac{n}{NT}\right] = \sum_{k=0}^{N-1} x(kT)e^{-(2\pi i/N)nk}. \quad (5)$$

The discrete S transform is obtained by allowing, $f \rightarrow n/NT$ and $\tau \rightarrow jT$

$$S\left[jT, \frac{n}{NT}\right] = \sum_{m=0}^{N-1} H\left(\frac{m+n}{NT}\right) e^{-2\pi^2 m^2/n^2} e^{i2\pi mj/N}, \quad (6)$$

where j, m and $n = 0, 1, \dots, N - 1$

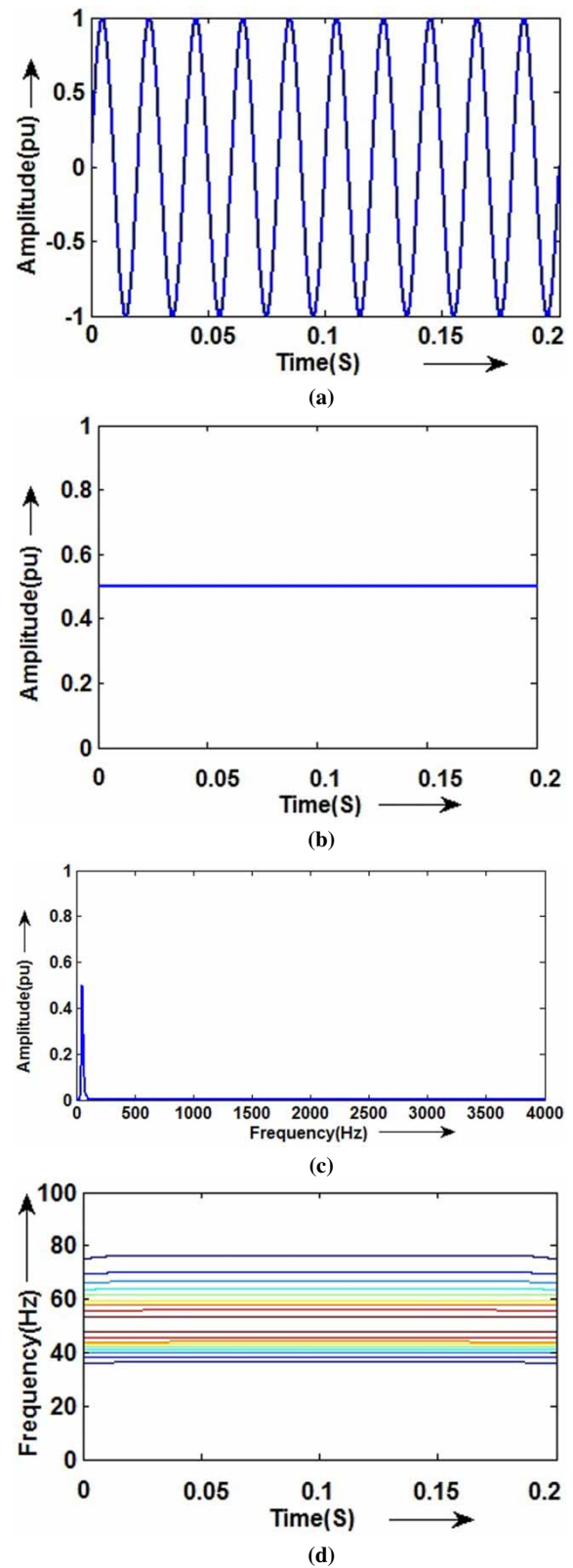


Figure 1. S-transform characteristics of normal waveform: (a) normal voltage signal, (b) time–maximum amplitude plot, (c) frequency–maximum amplitude plot and (d) time–frequency contour.

The discrete inverse of the S transform can be obtained as

$$x(kT) = \sum_{n=0}^{N-1} \left\{ \frac{1}{N} \sum_{j=0}^{N-1} S \left[jT, \frac{n}{NT} \right] \right\} e^{i2\pi nk/N}. \quad (7)$$

The rows and columns of the complex matrix obtained as output from S transform represent frequency and time, respectively. In this paper, the distinctive features belonging to each event signal have been extracted by applying S transform on the power

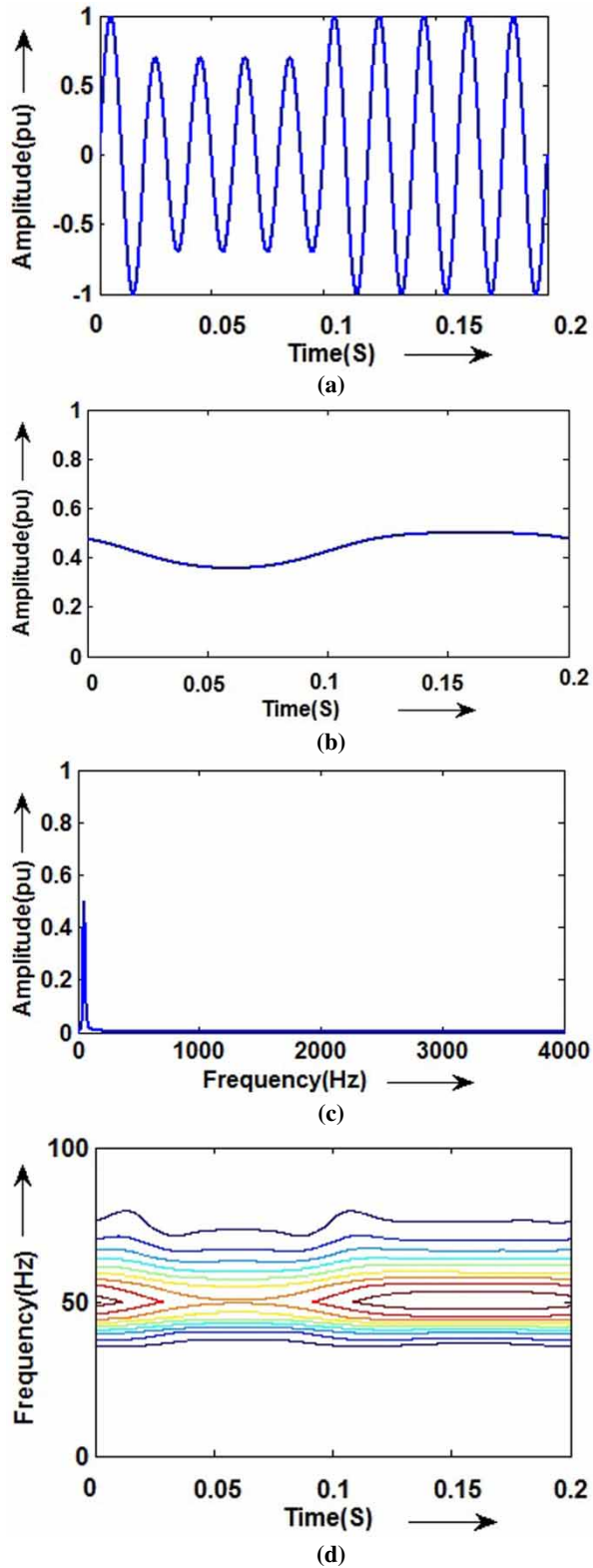


Figure 2. S-transform characteristics of Sag Waveform: (a) sag signal, (b) time–maximum amplitude plot, (c) frequency–maximum amplitude plot and (d) time–frequency contour.

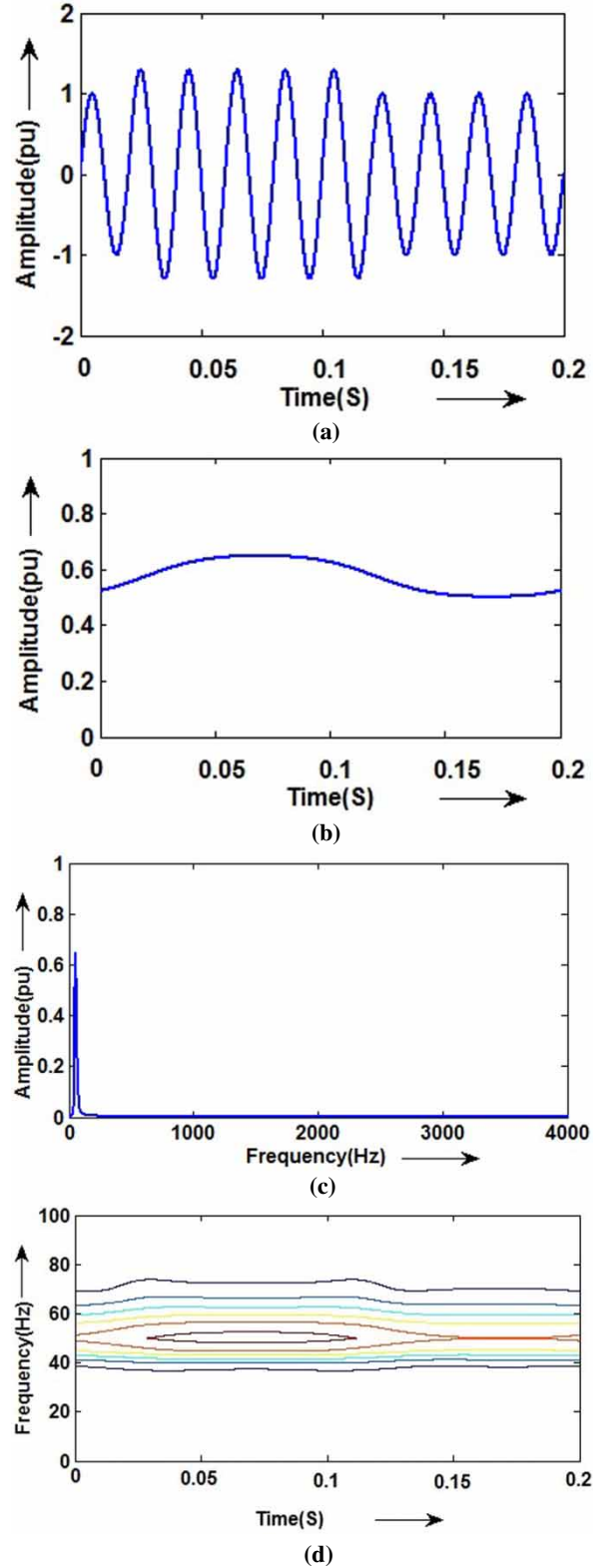


Figure 3. S-transform characteristics of swell waveform: (a) swell signal, (b) time–maximum amplitude plot, (c) frequency–maximum amplitude plot and (d) time–frequency contour.

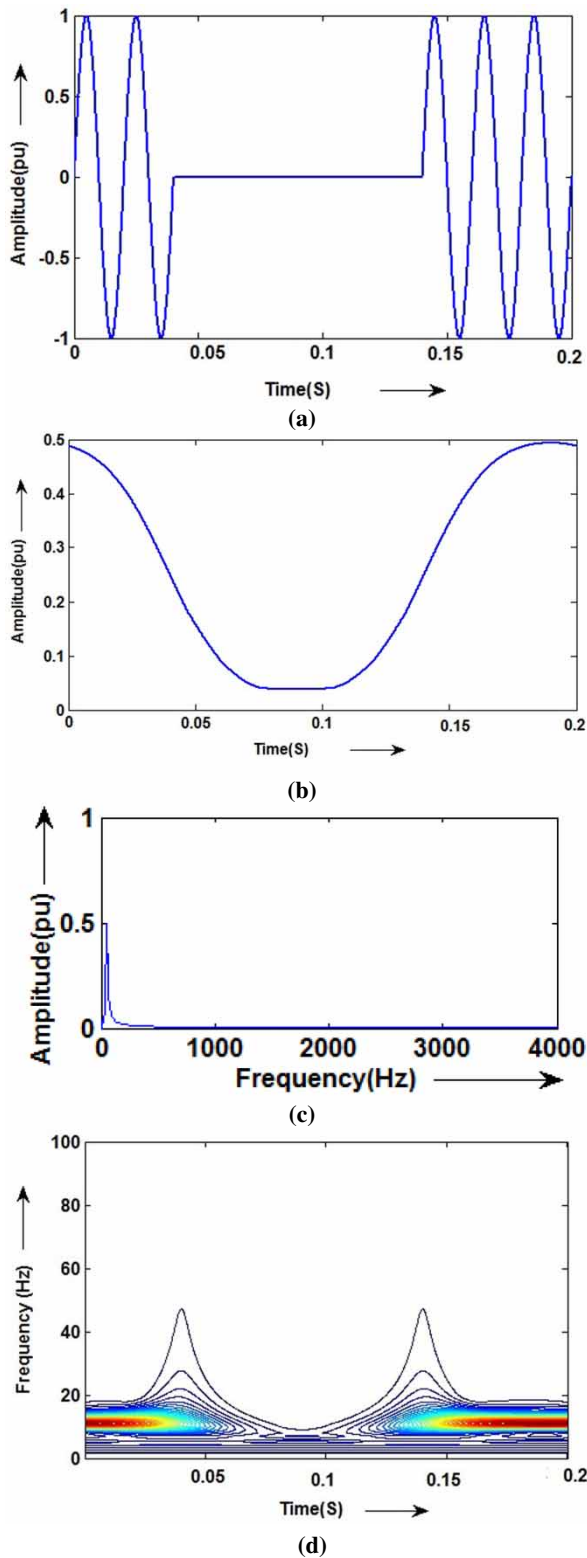


Figure 4. S-transform characteristics of interruption waveform: (a) interruption signal, (b) time–maximum amplitude plot, (c) frequency–maximum amplitude plot and (d) time–frequency contour.

quality event data. In comparison with the other time–frequency analysis methods, the S- transform has better noise immunity, which makes it an efficient way to recognize the power quality disturbances [28].

The simulated normal voltage signal (a) along with time–maximum amplitude plot (b), frequency–maximum

Table 1. Features for signal characterization.

Feature F1 =	Maximum of TmA-plot
Feature F2=	Minimum of TmA-plot
Feature F3=	Mean of TmA-plot
Feature F4=	Standard deviation of TmA-plot
Feature F5 =	abs (F1-F2-1/2)
Feature F6=	Standard deviation of the FmA-plot in the high-frequency area above 100 Hz
Feature F7=	Maximum amplitude of the FmA-plot in the high-frequency area (AHFMax)
Feature F8=	Minimum amplitude of the FmA-plot in the high-frequency area (AHFMax)
Feature F9=	F7-F8
Feature F10=	Kurtosis of the high-frequency area
Feature F11=	Skewness of the high-frequency area
Feature F12=	Mean of the FmA-plot
Feature F13=	Standard deviation of the FmA-plot
Feature F14=	Mean of the Fstd-plot
Feature F15=	Standard deviation of the Fstd-plot
Feature F16=	Standard deviation in the low-frequency area, below 100 Hz, of the Fstd-plot
Feature F17=	Standard deviation in the high-frequency area, above 100 Hz, of the Fstd-plot
Feature F18=	Total harmonic distortion
Feature F19=	Standard deviation of contour having the largest frequency amplitude of TF-contour
Feature F20=	Mean of contour having the largest frequency amplitude of TF-contour

amplitude plot (c) and time–frequency contour (d) is shown in Figure 1. Figures 2–4 illustrate these plots for sag, swell and interruption disturbances. The waveforms are modelled in MATLAB using the equations reported in [29].

In this work 20 features are extracted from the S-matrix for classification of power quality events and are tabulated in Table 1. By carefully analysing the characteristics of the different disturbances, it is evident that for proper classification, the features need to be extracted from time, frequency and amplitude, as well as from high- and low-frequency areas separately [29]. The classification accuracy depends on the number of features and the characteristics of the features obtained. The features have been obtained by applying standard statistical techniques on S-matrix.

3. ELM classifier

ELM, proposed by Huang et al. [30], is a single hidden layer feed forward network, where the input weights are chosen randomly and the output weights are calculated analytically. Sigmoid, sine, gaussian and hard limiting functions can be used as activation functions for the hidden neuron layer. The linear activation function is used for the output neurons [25]. ELM has several significant features, different from traditional popular gradient-based learning algorithms, for feed forward neural networks. The advantage of ELM includes faster learning speed, good generalization capability and evades issues like local minima [25].

Let us consider the ELM network with N observation samples $\{P_i, Q_i\}$, where $P_i = [p_{i1}, \dots, p_{in}] \in \mathcal{R}^n$ is an n -dimensional feature of the sample i and $Q_i =$

$[q_{i1}, \dots, q_{ic}] \in \mathfrak{N}^C$ is its coded class label. If the sample P_i is assigned to the class label c_k then the k th element of Q_i is one ($q_{ik} = 1$) and other elements are -1 . The output (\hat{Q}) of the ELM network with H hidden neurons and C distinct classes is defined as follows:

$$\hat{q}_{ik} = \sum_{j=1}^H U_{kj} G_j(W, B, P_i), \quad k = 1, 2, \dots, C, \quad (8)$$

where W represents $H \times n$ input weights, B represents $H \times 1$ bias of hidden neurons and U represents $C \times H$ output weights. $G_j(\cdot)$ is the output of the j th hidden neuron and is defined as follows:

$$G_j = G \left(\sum_{k=1}^n W_{jk} P_{ik} + b_j \right), \quad j = 1, 2, \dots, H, \quad (9)$$

where $G(\cdot)$ is the activation function.

In case of the radial basis function, the output of the j th neuron $G_j(\cdot)$ is defined as follows:

$$G_j = G(b_j \| P - W \|), \quad j = 1, 2, \dots, H, \quad (10)$$

where W and b_j are the centre and width of the radial basis function neuron.

In case of the sigmoidal activation function, the output of the j th neuron $G_j(\cdot)$ is defined as follows:

$$G_j = \tanh \left(b_j + \sum_{k=1}^n W_{jk} P_{ik} \right), \quad j = 1, 2, \dots, H. \quad (11)$$

Equation (8) can be written in matrix form as follows:

$$\hat{Q} = U Q_H, \quad (12)$$

where Q_H is the hidden layer output and is given as follows:

$$Q_H(W, B, P) = \begin{bmatrix} G_1(W, b_1, P_1) & G_1(W, b_1, P_2) & \dots & G_1(W, b_1, P_n) \\ \vdots & \vdots & \ddots & \vdots \\ G_H(W, b_H, P_1) & G_H(W, b_H, P_2) & \dots & G_H(W, b_H, P_n) \end{bmatrix}.$$

In the ELM algorithm, for a given number of hidden neurons, it is assumed that the input weights W and bias B of hidden neurons are selected randomly. By assuming the predicted output, \hat{Q} is equal to the coded labels Q , the output weights are estimated analytically as

$$\hat{U} = Q Q_H^+, \quad (13)$$

where Q_H^+ is the Moore–Penrose generalized pseudo-inverse of the hidden layer output matrix.

The ELM algorithm consists of the following steps:

1. Select the number of hidden neurons and a suitable activation function for a given problem.
2. Randomly, choose the input weight (W) and bias (B).

3. Analytically, calculate the output weight (U).
4. Use the calculated weights (W, U, B) for estimating the class label.

In this work the input to the ELM is the 20 features extracted from the S -matrix, while the output is the integer value of the class label.

4. Particle swarm optimization

In the recent years, PSO has been applied to solve a wide variety of engineering problems [31]. In PSO, particles constituting a swarm moves around the search space to determine the best solution [32]. Each particle in PSO adjusts its movement considering its own experience as well as the the experience of other particles.

The velocity of the i th particle in D -dimensional space is expressed as $V_i = (V_{i1}, \dots, V_{id}, \dots, V_{iD})$, the position of the i th particle is expressed as $X_i = (X_{i1}, \dots, X_{id}, \dots, X_{iD})$, the best position of the i th particle is expressed as $Pbest_i$ and the global best position among all the particles is expressed as $gbest$.

The velocity and position of the particles are updated using Equations (14) and (15)

$$V_{i,d}^{k+1} = \omega V_{i,d}^k + c_1 r_1 (X_{i,d}^k - Pbest_{i,d}^k) + c_2 r_2 (gbest_d^k - P_{i,d}^k), \quad (14)$$

$$X_{i,d}^{k+1} = X_{i,d}^k + V_{i,d}^{k+1} \quad (15)$$

where $1 \leq i \leq n, 1 \leq d \leq D$.

In the above expression, ω is the inertia weight, c_1 and c_2 are learning coefficients, r_1 and r_2 are separately generated uniformly distributed random numbers in the range $[0,1]$, k is the iteration number and n is the number of particles.

The following weighting function is usually used in Equation (14)

$$\omega = \omega_{max} - \left(\frac{\omega_{max} - \omega_{min}}{k_{max}} \right) \times k, \quad (16)$$

where ω_{max} and ω_{min} are the initial and the final weight, respectively, k is the current iteration number and k_{max} is maximum iteration number. The parameter ω regulates the trade-off between the global and the local exploration abilities of the swarm.

The first term of the velocity update equation (14) replicates the memory behaviour of the particles; the second term indicates the cognitive behaviour of the particles while the third part signifies the social behaviour of the particles. The convergence of the PSO algorithm is influenced by proper selection of parameters in the velocity update equation [33].

5. Enhanced PSO with craziness factor (crazy PSO)

To avoid being trapped into local optimum attempts have been made to improve the performance of PSO [34,35]. In this work, a craziness operator, along with chaotic weight updation is introduced in the conventional PSO algorithm to enhance the performance of PSO. In birds flocking or fish schooling, a bird or a fish frequently varies its direction. This phenomenon is introduced using a ‘‘craziness’’ factor [32]. To maintain the diversity of the particles, it is necessary to introduce craziness operation in a PSO algorithm [36]. This is done by updating the velocity of the particles using (17) and then applying the usual equations to modify the position and velocity using (14) and (15), respectively,

$$V_{i,n}^{k+1} = V_{i,n}^{k+1} + P(r_3) \times \text{sign}(r_3) \times V^{cr}, \quad (17)$$

where r_3 and V^{cr} are uniformly distributed random number uniformly distributed between 0 and 1.

$P(r_3)$ and $\text{sign}(r_3)$ are defined, respectively, as

$$\begin{aligned} P(r_3) &= 1 \text{ if } r_3 \leq p_{\text{craziness}} \\ &= 0 \text{ otherwise} \\ \text{sign}(r_3) &= 1 \text{ if } r_3 \geq 0.5 \\ &= -1 \text{ otherwise} \end{aligned}$$

where $p_{\text{craziness}}$ is a predefined probability of craziness.

The inertia weight is an important control parameter that affects the PSO’s convergence. The application of chaotic dynamics to update the inertia weight in PSO improves the searching behaviour and possibility of being trapped into local optimum is avoided [34,35]. In this context, the chaotic PSO approach based on the logistic equation applied for determining the weight factor. The logistic equation is defined as follows:

$$y(k) = \mu \times y(k-1) \times (1 - y(k-1)), \quad (18)$$

where k is the sample and μ is the control parameter, $0 < \mu \leq 4$. The behaviour of the system represented by (18) is greatly changed with the variation of μ . The value of μ determines, whether (18) stabilizes at a constant size, oscillates between limited sequences of sizes, or behaves chaotically in an unpredictable pattern. The behaviour of the system is sensitive to the initial value of ‘‘ y ’’ [37]. The system becomes deterministic, displaying chaotic dynamics when $\mu = 4$ and $y(1) \in \{0, 0.25, 0.5, 0.75, 1\}$. The parameter ‘‘ ω ’’ of (17) is modified by (19) through the following equation:

$$\omega_{\text{new}} = \omega \times y(k). \quad (19)$$

6. Crazy PSO-based ELM

In ELM, as the computation of output weights is based on the input weight and hidden biases, their choice

greatly influences the performance of ELM. The issues with random generation of parameters have been discussed in the literature [20]. In order to ensure better generalization ability of ELM, this paper proposes chaotic PSO algorithm-based selection of input weight and biases. The detailed steps of the proposed method are as follows:

- First, the swarm is randomly generated. Each particle in the swarm is composed of a set of input weights and hidden biases. All components in the particles are randomly initialized within the range of $[-1, 1]$.
- Second, for each particle, the corresponding output weights are computed. Then the fitness of each particle is evaluated. In this work the fitness of each particle is based on the root mean square error of the training set.
- Third, with the fitness of all the particles, the $Pbest_i$ and $gbest$ of the swarm are updated.
- Fourth, each particle updates its position and velocity according to (14) and (15), and a new population is generated.
- Finally, the above optimization process is repeated until maximum iteration.

Thus, the enhanced PSO algorithm is used to optimize the input weights and hidden biases.

7. Results and discussion

In order to investigate the performance of enhanced PSO-based ELM algorithm for classification of power quality disturbances, around 1100 samples of 8 classes of disturbances are generated using MATLAB software with each waveform based on the model as specified in reference [29]. As the electric power systems operate under noisy environment, the sampled signals come together with noise. The proposed approach must be analysed under noisy conditions. In this paper noisy signals are generated for all the 8 classes by adding different levels of gaussian white noise with signal-to-noise ratio of 30, 40 and 50 dB.

The proposed PSO ELM approach has been implemented in MATLAB and the percentage of classification accuracy is tabulated for training samples and testing samples in Tables 2 and 3. The results are presented for ELM structure with 100 hidden neurons and the sine activation function. The parameter settings for PSO are as follows: 50 particles, acceleration constants $c_1: 1.5, c_2: 1.5$, inertia weight $\omega_{\text{max}}: 0.9, \omega_{\text{min}}: 0.5$, and maximum iterations: 100. These parameters are selected carefully for efficient performance of the algorithms.

From the table, it is evident that the proposed approach has better classification accuracy in comparison to the conventional ELM approach and the PSO-based ELM approach.

Table 2. Training accuracy of different classification algorithms.

Signal	Without Noise			50 dB noise			40 dB noise			30 dB noise		
	Crazy	PSO-based	PSO-based	Crazy	PSO-based	PSO-based	Crazy	PSO-based	PSO-based	Crazy	PSO-based	PSO-based
	PSO-based	ELM	ELM	PSO-based	ELM	ELM	PSO-based	ELM	ELM	PSO-based	ELM	ELM
S1	99	98	99	99	99	95	100	98	96	96	84	91
S2	100	100	100	99	100	99	100	100	100	99	98	100
S3	100	100	100	100	100	100	100	100	100	100	100	100
S4	100	100	98	100	97	99	100	100	99	90	91	85
S5	100	100	100	99	98	99	99	99	100	99	99	99
S6	98	99	98	99	99	99	100	99	97	95	96	99
S7	100	100	100	99	100	100	100	100	100	81	79	81
S8	100	100	100	100	100	100	99	99	99	100	100	99

Table 3. Testing accuracy of different classification algorithms.

Signal	Without noise			50 dB noise			40 dB noise			30 dB noise		
	Crazy	PSO-based	PSO-based	Crazy	PSO-based	PSO-based	Crazy	PSO-based	PSO-based	Crazy	PSO-based	PSO-based
	PSO-based	ELM	ELM	PSO-based	ELM	ELM	PSO-based	ELM	ELM	PSO-based	ELM	ELM
S1	95	100	95	93	93	93	80	95	80	93	85	63
S2	100	100	100	100	100	100	100	100	98	77	73	81
S3	100	100	100	100	100	100	100	100	100	98	93	88
S4	100	97	100	100	97	100	100	94	100	97	94	61
S5	100	100	100	100	100	96	100	100	100	100	100	100
S6	100	100	100	100	96	100	100	100	100	100	96	96
S7	100	100	100	100	100	100	95	100	100	95	85	55
S8	100	100	100	98	100	100	90	98	98	98	100	94

Table 4. Classification accuracy comparison with other algorithms.

Accuracy	Crazy PSO-based ELM	Support vector machine	Back propagation neural network
Training(%)	98	97	98
Testing(%)	97	94	41

Further, it is evident that the classification accuracy of the proposed algorithm is superior to the other algorithms for the signals with disturbances.

The comparison with the support vector machine (SVM) and back propagation neural network presented in Table 4 illustrate the superiority of the proposed classifier. The results are presented by choosing the structure of the back propagation neural network similar to the ELM model, while in SVM, the radial basis function kernel is used as the kernel function, with softness factor and the kernel function parameters selected as 5 and 4, respectively.

8. Conclusion

This work envisages the application of craziness-based PSO to enhance the performance of ELM for feature classification of single and combined power quality disturbances. The input weight and hidden biases are adjusted using the optimization technique. The proposed technique is compared with PSO-based ELM and conventional ELM to highlight the classification accuracy. The results reveal that the proposed approach has superior classification accuracy under normal and noisy conditions. The fact that the proposed classifier results in better classification accuracy in comparison

with the SVM and back propagation neural network augurs well for its application in power quality event detection and classification. The addition of a regression network along with the proposed classifier to indicate the severity of the disturbance will help in initiating the corrective measures depending upon the nature and magnitude of the disturbance.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- [1] Axelberg PG, Gu IY, Bollen MH. Support vector machine for classification of voltage disturbances. *IEEE Trans Power Delivery*. 2007;22(3):1297–1303.
- [2] Monedero I, León C, Ropero J, et al. Classification of electrical disturbances in real time using neural networks. *IEEE Trans Power Delivery*. 2007;22(3):1288–1296.
- [3] Moravej Z, Banihashemi SA, Velayati MH. Power quality events classification and recognition using a novel support vector algorithm. *J Energy Conversion Management*. 2009;50:3071–3077.
- [4] Manimala K, Selvi K, Ahilaa R. Hybrid soft computing techniques for feature selection and parameter optimization in power quality data mining. *J Appl Soft Computing*. 2011;11:5485–5497.
- [5] Saini MK, Kapoor R. Classification of power quality events – a review. *J Electrical Power Energy Systems*. 2012;43:11–19.
- [6] Eris H, Yıldırım O, Eris_tı B, et al. Optimal feature selection for classification of the power quality events using wavelet transform and least squares support vector machines. *J Electrical Power Energy Systems*. 2013;49:95–103.

- [7] Deokar SA, Waghmare LM. Integrated DWT-FFT approach for detection and classification of power quality disturbances. *Electrical Power Energy Systems*. 2014;61:594–605.
- [8] Dehghani H, Vahidi B, Naghizadeh RA, et al. Power quality disturbance classification using a statistical and wavelet-based hidden Markov model with Dempster-Shafer algorithm. *J Electrical Power Energy Systems*. 2013;47:368–377.
- [9] Eris_tı H, Yıldırım O, Eris_tı B, et al. Optimal feature selection for classification of the power quality events using wavelet transform and least squares support vector machines. *J Electrical Power Energy Systems*. 2013;49:95–103.
- [10] Abdelsalam AA, Eldesouky AA, Sallamb AA. Characterization of power quality disturbances using hybrid technique of linear Kalman filter and fuzzy-expert system. *J Electric Power Systems Res*. 2012;83:41–50.
- [11] Rodríguez A, Aguado JA, Martín F, et al. Rule-based classification of power quality disturbances using S-transform. *Electric Power Systems Res*. 2012;86:113–121.
- [12] Huang N, Xu D, Liu X, et al. Power quality disturbances classification based on S-transform and probabilistic neural network. *J Neurocomput*. 2012;98:12–23.
- [13] Jayasree T, Devaraj D, Sukanesh R. Power quality disturbance classification using Hilbert transform and RBFnetworks. *J Neurocomput*. 2010;73:1451–1456.
- [14] Suja S, Jerome J. Pattern recognition of power signal disturbances using S transform and TT transform. *J Electrical Power Energy Systems*. 2010;32:37–53.
- [15] Kaewarsa S, Attakitmongcol K, Kulworawanichpong T. Recognition of power quality events by using multiwavelet-based neural networks. *J Electrical Power Energy Systems*. 2008;30:254–260.
- [16] Biswal B, Dash PK, Panigrahi BK. Power quality disturbance classification using fuzzy C-means algorithm and adaptive particle swarm optimization. *IEEE Trans Ind Electron*. 2009;56(1):212–220.
- [17] Zhu Q-Y, Qin AK, Suganthan PN, et al. Evolutionary extreme learning machine. *J Pattern Recognition*. 2005;38:1759–1763.
- [18] Suresh S, Saraswathi S, Sundararajan N. Performance enhancement of extreme learning machine for multi-category sparse data classification problems. *J Eng Appl Artif Intell*. 2010;23(7):1149–1157.
- [19] Lu S, Qiu X, Shi J, et al. A pathological brain detection system based on extreme learning machine optimized by Bat algorithm. *CNS Neurol Disord Drug Targets*. 2017;16(1):23–29.
- [20] Mohapatra P, Chakravarty S, Dash PK. An improved cuckoo search based extreme learning machine for medical data classification. *JSwarm and Evol Comput*. 2015;24:25–49.
- [21] Huang G-B, Chen L, Siew C-K. Universal approximation using incremental constructive feed forward networks with random hidden nodes. *IEEE Trans Neural Netw*. 2006;17(4):879–892.
- [22] Huang G, Chen L. Enhanced random search based incremental extreme learning machine. *J Neurocomput*. 2008;71(16–18):3460–3468.
- [23] Soria-Olivas E, Gomez-Sanchis J, Martin J, et al. BELM: Bayesian extreme learning machine. *IEEE Trans Neural Netw*. 2011;22(3):505–509.
- [24] Han F, Yao H-F, Ling Q-H. An improved evolutionary extreme learning machine based on particle swarm optimization. *J Neurocomput*. 2013;116:87–93.
- [25] Ahilaa R, Sadasivam V, Manimalaa K. An integrated PSO for parameter determination and feature selection of ELM and its application in classification of power system disturbances. *J Appl Soft Comput*. 2015;32:23–37.
- [26] Zhang Y-D, Zhao G, Sun J, et al. Smart pathological brain detection by synthetic minority oversampling technique, extreme learning machine and Jaya algorithm. *Multimed Tools Appl*. 2017: 1–20.
- [27] Stockwell RG, Mansinha L, Lowe RP. Localization of the complex spectrum: the S-transform. *IEEE Trans Signal Process*. 1996;44:998–1001.
- [28] Bhende CN, Mishra S, Panigrahi BK. Detection and classification of power quality disturbances using S-transform and modular neural network. *J Electric Power Systems Res*. 2008;78:122–128.
- [29] Huang N, Xu D, Liu X, et al. Power quality disturbances classification based on S-transform and probabilistic neural network. *J Neurocomput*. 2012;98:12–23.
- [30] Huang G-B, Zhu Q-Y, Siew C-K. Extreme learning machine: theory and applications. *J Neurocomput*. 2006;70(1–3):489–501.
- [31] Zhang Y, Wang S, Ji G. Review article: a comprehensive survey on particle swarm optimization algorithm and Its applications. *Math Probl Eng*. 2015.
- [32] Kennedy J, Eberhart R. Particle swarm optimization. *Proc IEEE Int Conf Neural Networks*. 1995: 1942–1948.
- [33] Clerc M, Kennedy J. The particle swarm-explosion stability and convergence in a multidimensional complex space. *IEEE Trans Evol Comput*. 2002;6(1):58–73.
- [34] Liu B, Wang L, Jin Y-H, et al. Improved particle swarm optimization combined with chaos. *JChaos, Solitons Fractals*. 2005;25(5):1261–1271.
- [35] Caponetto R, Fortuna L, FazziNo S, et al. Chaotic sequences to improve the performance of evolutionary algorithms. *IEEE Trans Evol Comput*. 2003;7(3):289–303.
- [36] Ho SL, Yang S, Ni G, et al. A particle swarm optimization-based method for multiobjective design optimizations. *IEEE Trans Magn*. 2005;41(5):1756–1759.
- [37] Song Y, Chen Z, Yuan Z. New chaotic PSO based neural network predictive control for Non linear process. *IEEE Trans Neural Netw*. 2007;18(2):595–600.