Fisher and Binary Particle Swarm Optimization Combination Approach to Feature Selection for Power System Stability Classification

N. N. Au[1,*], L. T. Nghia[1], V. V. Cuong[1], T. T. Giang[1], and T. V. Hien[1]

ARTICLE INFO

Article history:
Received: 20 October 2021
Revised: 21 December 2021
Accepted: 18 February 2022

Keywords:
Feature selection
Fisher
BPSO
Power System Stability
Classification.

ABSTRACT

Feature selection is an important task in data mining to reduce data dimension. Feature selection is a difficult problem, especially when the number of initial features is large. In classification of power system stability, feature selection is a challenging task due mainly to a large search space. This paper proposes fisher and binary particle swarm optimization (F&BPSO) combination approach to feature selection for stability of power system classification. The Fisher step, it is applied to find the features with high ranking. The output results of the fisher method are the input of BPSO. The BPSO step, BPSO now only has the task of finding the feature subset that reaches the fitness function value on the feature subset that have been reduced from the Fisher step. So, the approach makes the proposed method achieve the goal of reducing the maximum number of features, and having the highest accuracy. K-nearest neighbor (KNN, K=1) classifier is employed to evaluate the classification performance in the experiments on dataset. Test results on IEEE 39-bus diagram show that the proposed method achieves the goal of reducing features with high accuracy.

1. INTRODUCTION

Modern power systems suffer from operating pressure very close to a stable boundary limit, while power systems always face with stress condition that can easily cause vulnerability. Any faults in the power system will cause an unbalance between the mechanical power input to the generator and the electrical power output of the generator. As a result, the generator may lose synchronization with the power system and be automatically disconnected from the grid. Due to the very high non-linearity of the power system, traditional analytical methods take a lot of time to solve, causing delay in decision-making. Therefore, quickly detecting power system instability helps the control system to make timely decisions become the key factor to ensure stable operation of the power system [1], [2]. Classification method is one of the methods that can meet this requirement and has received great attention of researchers [3]–[6]. The key question in power system stability classification is whether the fault occurs, the power system is ''stable'' or ''unstable''.

The classification method has been used as an alternative to solving difficult problems that traditional methods of analysis cannot solve in terms of calculation [3]–[6]. By learning the database, the nonlinear input/output relationship between the power system operating parameters and stability can be quickly calculated [7]. However, if the classifier acts fast, the input feature subset must be the most important features. Therefore, the inputs need only be representative features, eliminate unnecessary and noisy features. Feature selection is not only important to reduce sensor measurement costs, but it also reduces the computational burden on the model.

In papers [8]–[10], the authors applied the ranking method to select variables, Fisher criterion is used in those. The paper [11] applied the ranking method thanks to the Relief algorithm to feature selection. So, the published works [8]–[11] mainly applied the ranking method to feature selection. This method evaluates each feature individually, without considering the context of a subset feature. so it can only provide local results. In previous studies [8]–[11] of feature selection for the classification of power system stability, the contributions of the Evolutionary computation techniques in feature selection are still limited. They are well-known for their global search ability, and have been applied to feature selection problems [12]. In which, BPSO is easier to implement, has fewer parameters, computationally less expensive. Due to these advantages, BPSO has been used as a promising method for feature selection problems. However, through experimentation, with a large number of features as power systems, it is difficult for BPSO to help reduce variables as deeply as desired. To overcome this obstacle, one idea proposed here is to combine the ranking method and BPSO for feature selection. With the support of the ranking method, the key features are selected first. The output result of selecting

[1]*HCM City University Technology and Education, Ho Chi Minh City, 100000, Vietnam.*
**Corresponding author:* N. N. Au; Email: aunn@hcmute.edu.vn.

features of the ranking method is the input of BPSO. This approach makes the proposed method achieve the goal of reducing the maximum number of features, and having the highest accuracy. Therefore, this paper proposes fisher and binary particle swarm optimization (F&BPSO) combination approach to feature selection for stability of power system classification.

The remainder of the paper is organized as follow: Background information is provided in Section 2. Section 3 offers the fisher and BPSO combination approach to feature selection. Section 4 presents experimental results with discussions. Section 5 provides conclusions.

BACKGRUOND

2.1 Fisher method

Fisher discrimination is based on fisher's linear discrimination function F(w) as projection from D-dimensional space onto a line in which manner the data is best separated. Given a set of n D-dimensional training samples $x_1, x_2, \ldots, x_n$ with n_1 samples in class 1 and n_2 samples in class 2, the task is to find the linear mapping, $y=w^T.x$, that maximizes F(w). This criterion evaluates the quality single variables. The quality of the variable is expressed through the value of F(w) as in equation (1). The value F is bigger means the feature is more important. Fisher is a criterion that was applied in many works with ranking method [8], [13]. By evaluating criterion of features as Equation (1), features are ranked by ordering the best of them and selecting for good features. The bigger feature F is the more important one.

$$F(w)=\frac{|m_1-m_2|^2}{\sigma_1^2+\sigma_2^2} \tag{1}$$

where m_i is the mean of n_i samples of class i and σ_i^2 is the variance of n_i samples of class i.

2.2 BPSO algorithm

PSO (Partical Swarm Optimization) is the optimal search algorithm proposed by Kennedy and Eberhart [14]. In the PSO, each candidate answer of the problem is encoded as an instance that moves through the search space. The whole herd seeks the optimal solution by updating the position of each individual based on their own experience and on neighboring individuals.

Generally, the vector $x_i = (x_{i1}, x_{i2}, ..., x_{iD})$ used in the PSO represents the position of the i^th^ instance. The vector $v_i = (v_{i1}, v_{i2},\ldots, v_{iD})$ is used in the PSO to rep-resent the velocity of the i th instance. D is the size of the search space. During the search, the best position for each previous individual was recorded as pbest. The best location of the herd is the gbest. The herd was randomly generated from the population. Finding the best solution by updating the velocity and position of each individual according to equations (2) and (3).

$$x_{id}^{t+1}=x_{id}^{t}+v_{id}^{t+1} \tag{2}$$

$$v_{id}^{t+1}=w*v_{id}^{t}+c_1r_{1i}(p_{id}-x_{id}^{t})+c_2r_{2i}(p_{gd}-x_{id}^{t}) \tag{3}$$

where, t is the t^{th} iteration of the search process. d is the size in the search space, $d\in D$. c_1, c_2 are acceleration constants. r_{1i}, r_{2i} are random values, valid in the range (0,1). p_{id} and p_{gd} represent pbest and gbest particles of size d. w is the inertial weight. v is the velocity, limited to the maximum velocity v_{max}, $v_{id}^t\in[-v_{max}, v_{max}]$.

The original PSO algorithm applied to the problem of continuity. Kennedy and Eberhart developed the BPSO algorithm for the discrete problem, Table 1 [15]. The velocity in BPSO represents the element that can take the value 1. Equation (2) is still used to update the velocity while x_{id}, p_{id} get the value 0 or 1. The function sigmoid $s(v_{id})$ is used to convert the value of vid into a range of values (0,1). The BPSO updates each instance's position using equations (4) and (5). The function rand() is a random function whose value is distributed in (0,1).

$$x_{id}=\begin{cases}1, if\ rand()<s(v_{id})\\ 0, otherwise\end{cases} \tag{4}$$

$$s(v_{id})=\frac{1}{1+e^{-v_{id}}} \tag{5}$$

Table 1. BPSO algorithm

BPSO algorithm
Begin Data set, k-fold; D: dimensionality of search space N : the population size; T : maximum iterations; c_1, c_2, v_{max}, w Randomly initialise the position and velocity of each particle; **while** *t ≤T* evaluate fitness of each particle according to Equation (6); **for** *i=1* **to** N update the *pbest* of particle *i*; update the *gbest* of particle *i*; **end** **for** *i=1* **to** *N* **for** *d=1* **to** *D* update the velocity of particle *i* according to Equation (2); update the position of particle *i* according to Equations (4) and (5); **end** **end** **end** calculate the classification error of the selected feature subset; return the position of *gbest* (the selected feature subset); return the classification error of the selected feature subset; **end**

2.3 Fitness function

In BPSO for feature selection, the number of selected variables and the classification error are evaluated during the execution of the algorithm. In order to achieve the dual goal of searching for the selected feature set with the least number of features and the smallest classification error, the paper proposes to apply the fitness function as Equation (6), which combines the two goals of minimizing the classification error rate and the number of features. This fitness function has the beauty of helping to strike a balance between wanting a deep reduction in the number of features while requiring the highest classification accuracy or the smallest classification error [12].

$$Fitness = \frac{Err}{TotalFeature - selectedFeature} \tag{6}$$

$$Acc(\%) = \frac{Accsample}{Totalsample} 100(\%) \tag{7}$$

$$Err(\%) = 1 - Acc(\%) \tag{8}$$

where, Fitness is the Fitness function; Acc is the classification accuracy of the selected feature subset; Accsample is the total number of samples the correct classification. Totalsample is the total number of samples; TotalFeature is the total number of features of the data set; selectedFeature is the number of selected features; Err is the classification error of the selected feature subset.

2. PROPOSED FISHER AND BPSO COMBINATION FEATURE SELECTION APPROACHE

To implement feature selection, we proposed that the feature selection process consists of the following three steps: Data generation, Fisher feature seletion approach, BPSO feature selection approach. In this paper, it is proposed to apply the 1-NN (K-NN, K=1) classifier to evaluate the classification accuracy. The 1-NN classifier is applied because of its fast computation and its simplicity. The prosed approach algorithm is named F&BPSO-1-NN.

Step 1. Data generation. This is the preparation step of initial data set for feature selection. Data set is noted by D(S,U). It formed by data set consists of stable samples, D(S), and un-stable samples, D(U).

Step 2. Fisher feature seletion. In this step, the fisher criterion is applied to rank the importance of the features in order to select the important features due to its simplicity. The calculated importance of the features is applied according to the formula (1). The output of this step is the input of step 3. It not only helps to select good potential candidate features, but also helps to reduce the computational burden for the next step.

Step 3. BPSO-1-NN feature selection approach. The input of this step is the output of step 2. This suggestion has an important meaning that it not only reduces the input variables but also helps to select the initial important input features for BPSO. BPSO algorithm is presented in Table 1. The fitness function drives the algorithm to find the best results with the smallest number of variables with the highest classification accuracy or the smallest error rate according to equations (6), (7), (8).

3. RESULTS AND DISCUSSION

4.1 Design of Experiment

In the experimental design for the power system stability classification, the initial feature selection analysis is very important because it directly affects the classification accuracy. This step defines a specific feature set that represents the database for learning of classifier. These initial features are the input variable representing the operating parameters of the power system and covering the operating status of the power system. The characteristic variable of the power system in transient mode or dynamic mode is the change in generator capacity, change of load capacity, change of power on transmission lines, and voltage drop at nodes,... right at the time of the faults. The obvious fact is that the change of active power and voltage drop contains very high information, and is strongly related to the stable power system state. Output features represent the stable conditions of the power system. Simulating observation results, if the relative angle of the generator rotors is larger than 180^0 then the system is 'Unstable', and less than 180^0 then the system is 'Stable'. The output variable is labeled binary. '0' is un-stable and '1' is stable.

The study was tested on the IEEE 39-bus scheme, Fig.1. It includes 39 buses, 19 loads, 10 generators. The diagram IEEE 39-bus scheme is well-known. It was used in many published works. The off-line simulation was implemented to collect data for training. Load levels are (80, 90, …, 120)% base load. The setting fault clearing time (FCT) is 50ms [16]. In this paper, all kinds of faults such as single phase to ground, double phase to ground, three phases to ground and phase-to-phase short-circuit are considered. Faults are tested in any buses and in each of 5% distances of long transmission lines of the test systems. For each of the considered load samples, the generator samples have been got accordingly by running optimal power flow (OPF) tool of Power-World software [16].

The input and output feature are x{$delV_{bus}$,$delP_{Load}$,$delP_{flow}$} and y{1,0}. Total of input features is 104, x{104(39+19+46)}. The symbol x{104(39+19+46) means the total number of variables is 104. Where 39 is 39 variables of delVbus , 19 is 19 variables of delPload, and 46 is 46 variables of delPflow. $delV_{bus}$ is the symbol for change voltage at nodes, $delP_{load}$ is the symbol for change of active power of load, $delP_{flow}$ is the symbol for change of active power on transmission lines. The number of output feature is one, y{1,0}. From simulating results, there are 1617 samples that include 834 stable samples and 783 un-stable samples. D(S,U) =1617, D(S)=834, and D(U)=783. The 1-NN is used as a classifier for evaluating accurate classification or error rate. Assessment of classification error is a cross-assessment method. It is a 10-fold cross-validation.

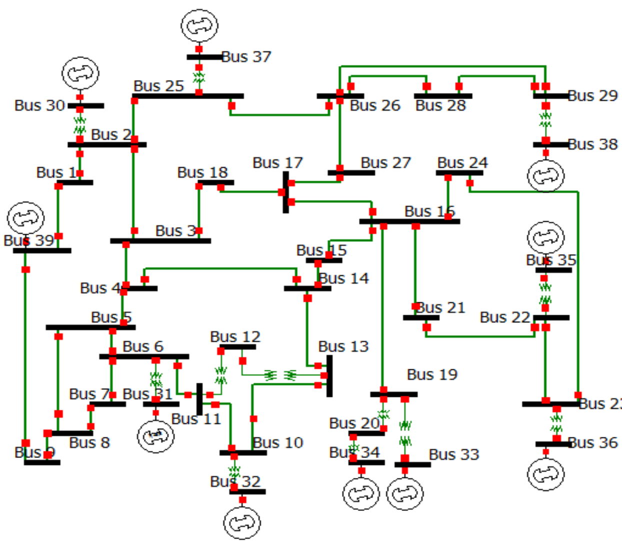

Fig.1. The IEEE 39-bus diagram.

4.2. Results

In this experiment, F&BPSO-1-NN is implemented to feature selection for stability data set of the IEEE 39-bus power system as described above. The results of feature selection of F&BPSO-1-NN are compared with the results of feature selection of the BPSO-1-NN.

BPSO works with different N values, namely 10, 20, 30, 40, and 50. The values of w is 0.9. The number of iterations is 100 for the program executions, T=100. The values of c_1 and c_2 are selected unchanged during program execution, c_1=2, c_2=2. The program is executed on Matlab 2018a software.

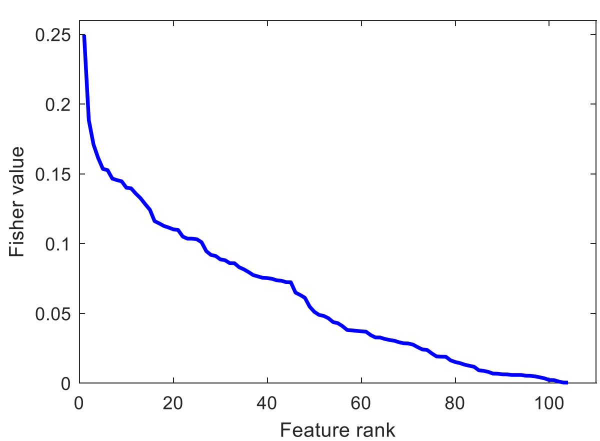

Fig. 2 Fisher value of features.

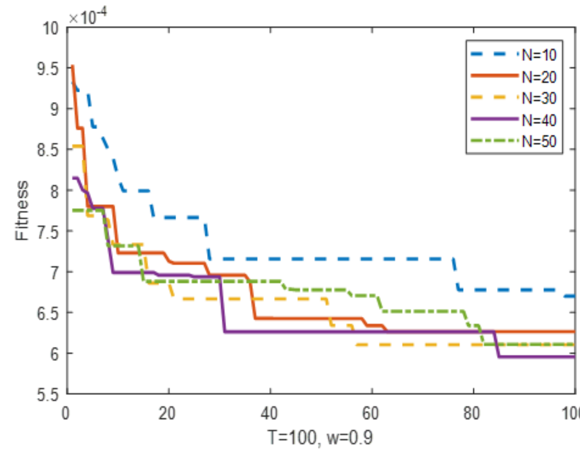

Fig. 3 Convergence characteristics of BPSO-1-NN, d=104.

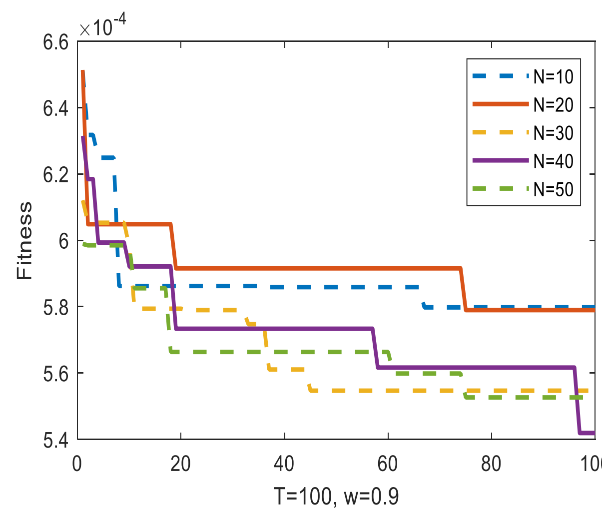

Fig. 4 Convergence characteristics of F&BPSO-1-NN, d=20.

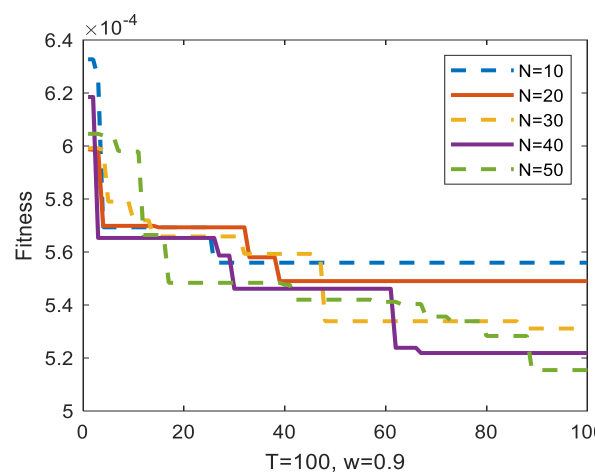

Fig. 5 Convergence characteristics of F&BPSO-1-NN, d=30.

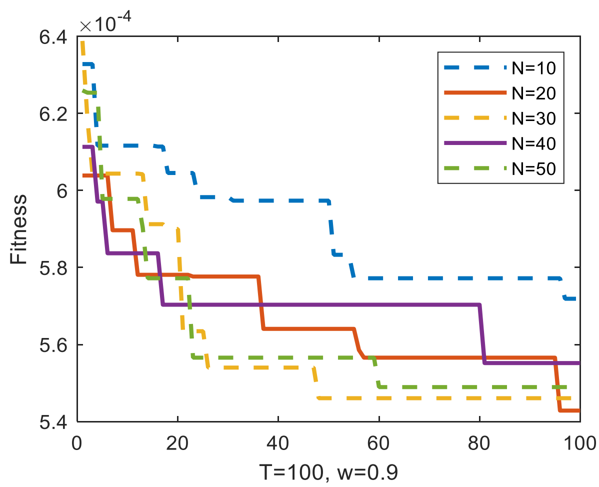

Fig. 6 Convergence characteristics of F&BPSO-1-NN, d=40.

Table 2. Results of F&BPSO-1-NN and BPSO-1-NN

Algorithm	d	N	nf	Best Fitness	Err(%)
F&BPSO-1-NN	20	10	8	5.7978e-04	6.18
		20	10	5.7895e-04	6.12
		30	7	5.5467e-04	5.88
		40	7	5.1492e-04	6.00
		50	10	5.5264e-04	5.69
	30	10	15	5.5589e-04	5.5
		20	15	5.4894e-04	5.63
		30	12	5.3100e-04	5.88
		40	8	5.2180e-04	5.32
		50	8	5.1536e-04	5.38
	40	10	11	5.7188e-04	5.63
		20	14	5.4284e-04	5.50
		30	10	5.4606e-04	5.69
		40	16	5.5518e-04	5.44
		50	15	5.4894e-04	5.13
BPSO-1-NN	104	10	32	6.6996e-04	5.44
		20	27	6.2646e-04	5.38
		30	27	6.1040e-04	5.13
		40	23	5.9552e-04	5.13
		50	23	6.1079e-04	5.07

Step 2. Fisher feature selection. Applying Fisher criterion as equation (1), the Fisher values of all the features are assessed and shown in Fig. 2 in descending order.

Step 3. BPSO-1-NN feature seletion approach. BPSO-1-NN implements feature selection with an initial number of input variables (d) of 104. The convergence characteristics are shown in Fig. 3 and the results of feature selection are shown in Table 2.

From step 2, the number of input selected variables for step 3 are 20, 30, and 40. The convergence characteristics of the algorithm are shown in Fig. 4, Fig. 5, and Fig. 6. the results of feature selection are shown in Table 2. In Table 2, nf is the symbol for the number of selected variables.

3.3. Discussion

According to Table 2, with the initial number of input primitive variables of 104, implementing the BPSO-1-NN algorithm has reduced the number of variables to 23 variables with the classification error of 5.07% or the classification accuracy of 94.93%.

Also according to Table 2, the number of input variables for step 3 is selected as 20, 30, 40. In step 3, with 20 input variables, the proposed algorithm achieves 7 variables with classification error of 5.88% or classification accuracy of 94.12%. With 30 input variables, the proposed algorithm achieves 8 variables with classification error of 5.32% or classification accuracy of 94.68%. With 40 input variables, the proposed algorithm achieves 10 variables with classification error of 5.69% or classification accuracy of 94.31%.

Compared with the BPSO-1-NN algorithm, in experimental cases, the proposed algorithm F&BPSO-1-NN is capable of reducing the number of variables to 69.5%, 65.2%, 56.5%, while the classification accuracy is only decreased by 0.78%, 0.22%, and 0.59% respectively. The classification accuracy for all cases is more than 94%. This is also the accepted result in previously published works [17]–[20].

4. CONCLUSIONS

The paper has introduced the new method to feature selection for stability power systems classification. The procedures proposed in the algorithm are specific, clear and very promising to be applied in selecting variables for evaluating the stability of the power system.

The test results on the IEEE 39-bus diagram show that the variable reduction algorithm is very effective. Compared with the original 104 variables, the proposed algorithm has ability to reduce the number of variables to 7 variables. The rate of variables is reduced to about 93.2%. This has great significance in reducing sensor measurement costs, reducing computational costs for recognition models. The results of the study contribute to enriching the research direction of power system stability by classification method.

ACKNOWLEDGEMENTS

This work belongs to the project in T2021-48TĐ funded by Ho Chi Minh City University of Technology and Education, Vietnam.

REFERENCES

[1] B. Sun, M. Liu, L. Zhu, N. Liu, X. Qiu, and Z. Zhuang, "Emergency Control Strategy Based on Multi-agent Theory under Blackout," *Energy Power Eng.*, vol. 05, no. 04, pp. 717–721, 2013, doi: 10.4236/epe.2013.54B139.

[2] Y. V. Makarov, V. I. Reshetov, V. A. Stroev, and N. I. Voropai, "Blackout prevention in the United States, Europe, and Russia," *Proc. IEEE*, vol. 93, no. 11, pp. 1942–1954, 2005, doi: 10.1109/JPROC.2005.857486.

[3] A. Sharifian and S. Sharifian, "A new power system transient stability assessment method based on Type-2 fuzzy neural network estimation," *Int. J. Electr. Power Energy Syst.*, vol. 64, pp. 71–87, 2015, doi: 10.1016/j.ijepes.2014.07.007.

[4] W. D. Oliveira, J. P. A. Vieira, U. H. Bezerra, D. A. Martins, and B. das G. Rodrigues, "Power system security assessment for multiple contingencies using multiway decision tree," *Electr. Power Syst. Res.*, vol. 148, pp. 264–272, 2017, doi: 10.1016/j.epsr.2017.03.029.

[5] S. M. Ashraf, A. Gupta, D. K. Choudhary, and S. Chakrabarti, "Voltage stability monitoring of power systems using reduced network and artificial neural network," *Int. J. Electr. Power Energy Syst.*, vol. 87, pp. 43–51, 2017, doi: 10.1016/j.ijepes.2016.11.008.

[6] J. Mitra, M. Benidris, N. Nguyen, and S. Deb, "A Visualization Tool for Real-Time Dynamic Contingency Screening and Remedial Actions," *IEEE Trans. Ind. Appl.*, vol. 53, no. 4, pp. 3268–3278, 2017, doi: 10.1109/TIA.2017.2686353.

[7] R. Zhang, Y. Xu, Z. Y. Dong, K. Meng, and Z. Xu, "Intelligent systems for power system dynamic security assessment: Review and classification," *2011 4th Int. Conf. Electr. Util. Deregul. Restruct. Power Technol.*, pp. 134–139, 2011, doi: 10.1109/DRPT.2011.5993876.

[8] Z. Y. Dong, Z. Rui, and Y. Xu, "Feature selection for intelligent stability assessment of power systems," in *2012 IEEE Power and Energy Society General Meeting*, 2012, pp. 1–7, doi: 10.1109/PESGM.2012.6344780.

[9] K. S. Swarup, "Artificial neural network using pattern recognition for security assessment and analysis," *Neurocomputing*, vol. 71, no. 4–6, pp. 983–998, 2008, doi: 10.1016/j.neucom.2007.02.017.

[10] Y. Xu, Z. Y. Dong, K. Meng, R. Zhang, and K. P. Wong, "Real-time transient stability assessment model using extreme learning machine," *IET Gener. Transm. Distrib.*, vol. 5, no. 3, p. 314, 2011, doi: 10.1049/iet-gtd.2010.0355.

[11] Y. Xu, S. Member, Z. Y. Dong, S. Member, L. Guan, and R. Zhang, "Preventive Dynamic Security Control of Power Systems Based on Pattern Discovery Technique," vol. 27, no. 3, pp. 1236–1244, 2012.

[12] B. Xue, M. Zhang, and W. N. Browne, "New fitness functions in binary particle swarm optimisation for feature selection," *2012 IEEE Congr. Evol. Comput. CEC 2012*, 2012, doi: 10.1109/CEC.2012.6256617.

[13] A. R. Webb and K. D. Copsey, *Statistical Pattern Recognition*. 2011.

[14] J. Kennedy and R. Eberhart, "Particle Swarm Optimization," *IEEE Int. Conf. , Perth, WA, Aust.*, no. ISBN: 0-7803-2768-3, pp. 1942–1948, 1995.

[15] J. Kennedy and R. C. Eberhart, "Discrete binary version of the particle swarm algorithm," *Proc. IEEE Int. Conf. Syst. Man Cybern.*, vol. 5, pp. 4104–4108, 1997, doi: 10.1109/icsmc.1997.637339.

[16] J. D. Glover, M. S. Sarma, and T. J. Overbye, *Power system analysis and design*, Fifth Edit. Christopher M. Shortt, 2012.

[17] N. Amjady and S. A. Banihashemi, "Transient stability prediction of power systems by a new synchronism status index and hybrid classifier," *IET Gener. Transm. Distrib.*, vol. 4, no. 4, p. 509, 2010, doi: 10.1049/iet-gtd.2009.0255.

[18] S. Kalyani and K. S. Swarup, "Pattern analysis and classification for security evaluation in power networks," *Int. J. Electr. Power Energy Syst.*, vol. 44, no. 1, pp. 547–560, 2013, doi: 10.1016/j.ijepes.2012.07.065.

[19] A. Karami and S. Z. Esmaili, "Transient stability assessment of power systems described with detailed models using neural networks," *Int. J. Electr. Power Energy Syst.*, vol. 45, no. 1, pp. 279–292, 2013, doi: 10.1016/j.ijepes.2012.08.071.

[20] A. Y. Abdelaziz and M. A. El-Dessouki, "Transient Stability Assessment using Decision Trees and Fuzzy Logic Techniques," *Int. J. Intell. Syst. Appl.*, vol. 5, no. 10, pp. 1–10, 2013, doi: 10.5815/ijisa.2013.10.01.