Multi-Goal Feature Selection Function in Binary Particle Swarm Optimization for Power System Stability Classification

Ngoc Au Nguyen
Ho Chi Minh City University of Technology and Education, Vietnam
aunn@hcmute.edu.vn

Hoang Minh Vu Nguyen
Ho Chi Minh City University of Architecture, Vietnam
vu.nguyenhoangminh@uah.edu.vn

Trong Nghia Le
Ho Chi Minh City University of Technology and Education, Vietnam
trongnghia@hcmute.edu.vn
(corresponding author)

Received: 20 February 2023 | Revised: 3 March 2023 | Accepted: 8 March 2023

ABSTRACT

The input feature is vital for power system stability classification. Feature selection can reduce the size of the input feature, making classifier training easier, and the small size of the input feature subset also reduces the cost of purchasing sensor measurement equipment. Therefore, feature selection works require an efficient seeking method. Binary Particle Swarm Optimization (BPSO) is a simple and easy-to-implement evolutionary calculation technique. While the classification accuracy of BPSO is essential, it is also needed to drive the algorithm to minimize the number of variables. The proposed approach is to apply a multi-objective function to help the BPSO algorithm identify the subset of features that can achieve the minimum number of variables and the highest classification accuracy. The k-nearest neighbor classifier is employed in the experiments to evaluate the classification performance on the IEEE 39-bus dataset.

Keywords-feature selection; classification; power system stability; BPSO; K-nearest neighbor classifier

I. INTRODUCTION

Stability analysis of electrical systems is a highly complex problem [1]. Stability losses of electrical system cause widespread power outages and often significant economic damage [2]. The load growth is outpacing the investment in transmission lines and power generation. As a result, incidents can easily lead to power system instability and outages [3, 4]. Therefore, rapid detection of power system instability is crucial [1], but traditional analytical methods cannot achieve it. This difficult requirement can be resolved using a recognition method [5-7]. Once trained on input-output data, operating modes, and stability modes of the electrical system, the recognition system will quickly detect the instability of the electrical system [8]. However, the abundance of input data is a significant obstacle for the recognition system in learning and operation, adding a high cost of measuring sensor data and collecting data. Therefore, in recognition problems, the important issue is to make the input variables accurate and straightforward. In recognition of the stability analysis of electrical systems, previous studies have proposed solutions for the primary variable selection process, mainly using ranking methods to select variables. The Relief method and the divergence criterion are applied in [9]. In [9-10], the Fisher criterion is applied. Ranking methods are based on statistical criteria but do not consider the recognition error and are not optimal. Advanced calculation techniques can help optimally expand the search space. Binary Particle Swarm Optimization (BPSO) is a simple and easy-to-implement candidate computational optimization algorithm for feature selection in the stable electrical system classification problem. In this, the important objective function is the classification error. However, more than one objective functions based solely on classification error are required to guide the BPSO algorithm toward selecting small and low-dimensional inputs [11]. This paper proposes the application of a multi-objective feature selection function based on the minimum classification error with minimum size of feature-subset or the highest classification accuracy with the least number of input feature-subset. The K-Nearest Neighbor (K-NN) algorithm is recommended for assessing the classification performance in the BPSO algorithm. The proposed method is tested on the IEEE-39 bus power system data.

II. BACKGROUND

A. Goal Function

1) Correct Classification

In the steady-state power system classification, the Correct Classification is calculated according to (1):

$$CC(\%) = \frac{CSC}{TSC} 100(\%) \tag{1}$$

where CSC (Correct Sample Classification) is the number of correctly classified samples and TSC (Total Sample Classification) is the total number of samples in the sample set.

2) Error Classification

Error Classification is calculated by:

$$EC(\%) = 1 - CC(\%) \tag{2}$$

3) Multi-Objective Selection Function

In BPSO (Figure 1), the total number of variables in the data set is the size of the search space. A binary string represents each candidate (a particle). The bits '1' and '0' correspond to the selected and unselected variable, respectively.

```
Begin
    Data set, k-fold; D: dimensionality of search space
    N : the population size; T : maximum iterations;
  c1, c2, vmax, w
  Randomly initialize the position and velocity of each particle;
   while t ≤T
    evaluate fitness of each particle according to (6);
      for i=1 to N
      update the p_best of particle i;
      update the g_best of particle i;
       end
     for i=1 to N
        for d=1 to D
        update the velocity of particle i according to (2);
        update the position of particle i according to (4) and
        (5);
        end
      end
    end
  calculate the classification error of the selected feature subset;
  return the position of g_best (the selected feature subset);
  return the classification error of the selected feature subset;
end
```

Fig. 1. The BPSO algorithm.

The single objective function (2) only contains the classification error, which can lead the BPSO algorithm to converge to a solution with many variables. There are two main objectives in variable selection: the lowest classification error and the smallest number of selected variables. Therefore, the variable selection objective function (3) combines these two objectives. In practice, the number of selected variables (SF) is larger than the classification error (EC). To balance these two quantities, the second quantity of the objective function is the ratio of the number of selected variables to the total number of variables, which is aimed at converting the second quantity into the sequence (0, 1].

$$\mathrm{GF} = EC + \frac{SF}{TF} \tag{3}$$

The objective function (3) reflects the importance of the misclassification error and the minimum number of variables having equal weighting or a weighting ratio of 1:1. If we want to consider the importance of the classification error rate and the number of selected variables differently, the α weight is introduced into (3) to become:

$$\mathrm{GF} = \alpha.EC + (1-\alpha).\frac{SF}{TF} \tag{4}$$

where $\alpha \in [0,1]$. For $\alpha \in (0.5,1]$, the classification error has a higher weight than the number of selected variables, while the opposite takes place when $\alpha \in (0,0.5)$ and when $\alpha = 0.5$ the importance of the classification error and the number of selected variables is equal.

B. BPSO Algorithm

The PSO algorithm is a nature-inspired approach proposed for solving continuous optimization problems [12]. The BPSO algorithm was developed for discrete optimization problems [13]. A matrix representing the stability of an electric system is presented below. This data matrix is represented by the vector x_i. The vector $x_i = (x_{i1}, x_{i2}, \ldots, x_{iD})$ used in the BPSO represents the position of the i^{th} sample. The vector $v_i = (v_{i1}, v_{i2}, \ldots, v_{iD})$ is used to represent the velocity of the i^{th} sample. The size of the search space is designated as D. Throughout the search process, the optimal position of each individual is recorded as p_{best}. The global best position, representing the most advantageous location for the entire herd, is referred to as g_{best}. The individuals in the herd are randomly generated from the population. The objective is to find the optimal solution by iteratively updating the position and velocity of each individual based on (5) and (6).

$$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t+1} \tag{5}$$

$$\begin{aligned} v_{id}^{t+1} = {} & w * v_{id}^{t} + c_1 r_{1i}(p_{id} - x_{id}^{t}) \\ & + c_2 r_{2i}(p_{gd} - x_{id}^{t}) \end{aligned} \tag{6}$$

where t is the t^{th} iteration of the search process, $d \in$D is the size in the search space, 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 p_{best} and g_{best} particles of size d, w is the inertial weight, v is the velocity, limited to the maximum velocity v_{max}, and $v_{idt} \in [-v_{max}, v_{max}]$.

The velocity in BPSO represents the element that can take the value 1. Equation (6) is still used to update the velocity while x_{id}, p_{id} get the value 0 or 1. The sigmoid function $s(v_{id})$

is used to convert the value of v_{id} into a range of values in (0,1). The BPSO updates each sample's position using (7) and (8). The rand() function is a random number generator with a uniformly distributed output between 0 and 1.

$$x_{id} = \begin{cases} 1, if\ rand() < s(v_{id}) \\ 0, otherwise \end{cases} \tag{7}$$

$$s(v_{id}) = \frac{1}{1+e^{-v_{id}}} \tag{8}$$

III. STEPS TO IMPLEMENT VARIABLE SELECTION IN THE STABILITY CLASSIFICATION PROBLEM OF THE POWER SYSTEM.

The methodology for selecting variables in the stability classification of power systems involves the steps depicted in Figure 2: Initial variable selection analysis, data collection, the proposed variable selection method, and, lastly, the implementation and results.

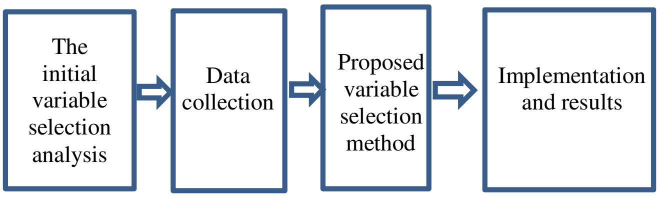

Fig. 2. Steps to implement variable selection.

Step 1: Initial variable selection analysis. Analysis for system stability assessment found that major disturbance events, such as short circuits, have a high potential to cause system instability. Thus, short circuits are of particular concern. Examining these input variables is crucial for the determination of their impact on the stability of the electrical system. This allows for the selection of variables that are strongly correlated with system stability, and facilitates the identification of key initial variables while eliminating unnecessary ones. The goal of variable selection is to minimize the number of variables used in the analysis, while maintaining the highest level of accuracy in the classification process. The selection of initial variables is a preliminary screening step to make the subsequent steps easier. The selected variables must have a low cost of sensor measurement. Suppose all variables related to the electrical system are considered. In that case, the number of variables is very large, such as the powerful effect of branches, the power resistance of branches, the powerful effect of generators, the power resistance of generators, generator voltage current, generator speed, the relative angle deviation of the generators from the standard machine, etc. The analysis of the electrical system's stability reveals a direct correlation between the loss of stability and changes in active power and voltage. Therefore, it is imperative to consider these variables.

Step 2: Collect and organize data. The initial data collected can be historical data stored in the data collection system when operating the power system, or they can be simulated using specialized software to build the initial data set. The data consist of input and output data. The data set is stored as a matrix of data. The input data includes data columns containing information about variables, and the data matrix rows contain the variables' parameters. The output data is a matrix containing information on whether the system is stable or unstable, assigned binary labels, matrix y (bit '1' is stable, and bit '0' is unstable). Figure 3 presents the organization of the input and output data matrix.

$$\begin{array}{cc} & \text{Variation} \\ \text{Sample} & \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ x_{n1} & x_{n2} & \cdots & x_{nd} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} \end{array}$$

Fig. 3. Input and output matrix.

Step 3: Proposed variable selection method. As presented above, studies on stability recognition of electrical systems mainly apply statistical standards to rank variables. The result is that the ranking method still does not give the result of how many variables are selected and what they are. The automatic and optimal selection of variables is still a big question. The BPSO algorithm is a promising candidate for variable selection in the stability classification of electrical systems and the use of one multi-objective function to guide the BPSO algorithm to find the best results is proposed. The multi-objective function is introduced in Section II.A.3. When selecting recognition sets to evaluate classification accuracy, they must meet the requirements of simplicity without considering computational time. Through our experiment, we found that the K-NN tool meets this criterion.

Step 4: Implementation and results. Choosing the optimal set of variables is a challenging problem as it depends on too many factors. Thus, many aspects must be considered to choose the best possible set of variables. Therefore, the multi-objective function (4) is suggested to be applied for variable selection in the implementation step.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Results

The experiment was conducted on the IEEE 39 bus system, which comprises 39 buses, 10 of which are dedicated to generation. Additionally, the system consists of 12 transformers, 10 generators, 34 transmission lines, and 19 loads. Ten generators are connected from bus 30 to bus 39, where bus 39 is chosen as the reference bus, 9 buses are PV buses, and the remaining 29 buses are PQ buses, with two different voltage levels of 345kV and 20kV. The IEEE 39 bus system diagram is shown in Figure 4. A HP Probook Laptop, Intel(R) Core(TM) i7 8th Gen, 8GB RAM was used.

Initial variable selection analysis: The input variables are in the dynamic state containing information about the established operating mode and the fault mode, combined with expert opinions to select the initial variables, including voltage deviation at the buses {delVbus}, load power deviation {delPload}, and transmission line power deviation {delPflow}. The input and output variables are x and y. The survey on the diagram has 39 voltage deviation variables at the buses, 19 load power effect variables, and 46 branch power effect variables. The total number of input variables is 104 and there is only 1 output variable.

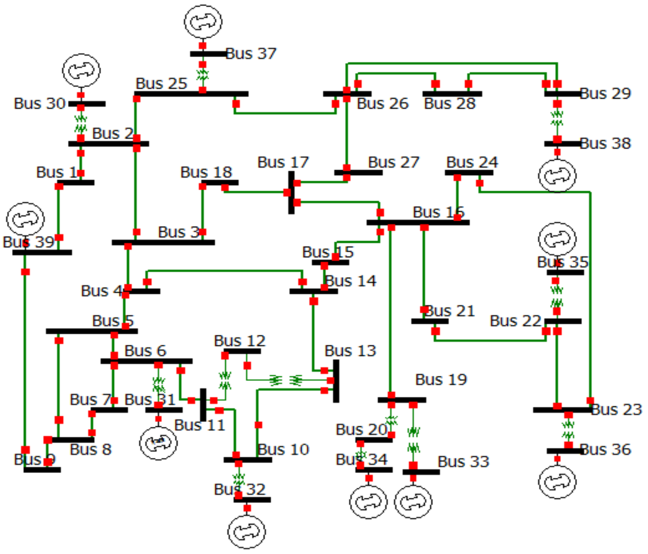

FIG. 4. IEEE 39 bus system diagram.

Data collection: Offline simulation was conducted using Powerworld 18 software to collect data for power system stability assessment. In this experiment, the load level considered is 100% of the base load, and the fault scenario is a short circuit. In the load flow examination, the software computes the optimal power distribution and the corresponding power flow. The faults considered are three-phase short circuits, one phase to ground, and two phases at all busbars, and along all transmission lines, with each spacing of 5% of the length of the transmission line. The cutting time of the short circuit is 50ms [14]. The state of the power system is stable when the rotor angle deviation of any two generators does not exceed 180^o, and it is not stable when it does. The simulation results show that the data set contains 317 samples, 177 stable and 140 unstable.

BPSO Algorithm and multi-objective variable selection: The K-NN (K=1, 1-NN) classifier is selected as the partitioning method. The accuracy of 1-NN classification was evaluated with the cross-validation method with k-folds=10. It involves dividing the dataset into 10 equal parts or "folds." The model is then trained on 9 of the folds and tested on the remaining 1 fold. This process is repeated 10 times, each time using a different fold for testing, while the other 9 folds are used for training. The result is the average of these 10 times. The calculation functions are supported by Matlab 2018b. The BPSO-1-NN algorithm is applied to execute the variable selection. The considered values to evaluate the importance of components in the multi-objective optimization variable selection function (4) are: N = {10, 20, 30, 40, 50}, w = 0.9, T = 100, c_1 = 2, c_2 = 2, while α is considered to have values in (0,1] with a step of 0.1. The convergence trajectory of the algorithm is illustrated in Figure 5. The results of the variable selection, objective function value, and classification accuracy are presented in Table I.

B. Discussion

The algorithm selected 11 variables with corresponding values of α weight of 0, 0.2, 0.3, and 0.4. Among them, α=0.2, N=20, and SF=11 achieved 95.6% accuracy in the testing phase.

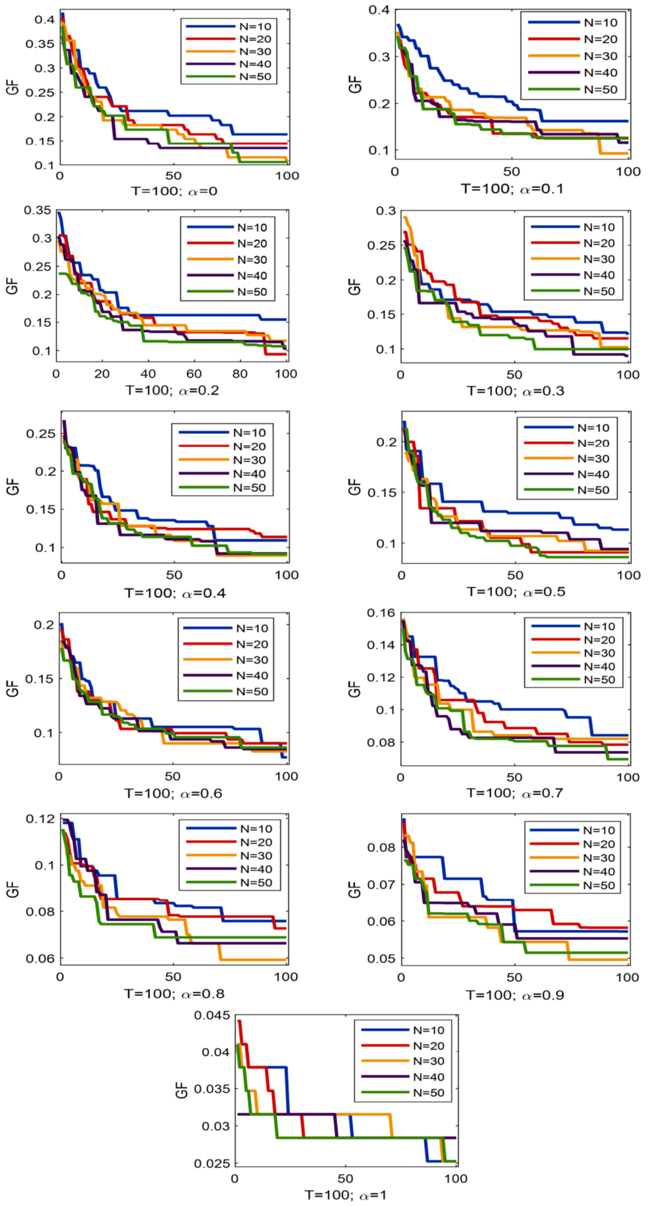

Fig. 5. The convergence characteristics of the algorithm.

V. RESULTS

The accuracy when α=0.8, *N*=30, and SF=19 was 97.1%. The accuracy when α=0.9, *N*=50, and SF=24 was 96.8%. The accuracy when α=1, *N*=20, and SF=46 was 97.1%. This shows that as the value of α in the objective function (4) increases, the importance of the classification error decreases and the accuracy of the classification increases, but with a larger number of variables. In the case of α=1, the objective function (4) becomes a single objective, and only the single objective of the classification error remains. There, 46 variables were selected, with the number of selected variables decreasing by 55.7%, while in the case of α=0.2, only 11 variables were selected, with the number of selected variables decreasing by 89.4%. In terms of classification accuracy, the case of 11 variables had accuracy only 1.5% lower than that of the case of 46 variables, while the number of selected variables decreased by 33.7%. A classification accuracy of 95.6% is considered acceptable for stable power system classification [15–17].

TABLE I. RESULTS OF THE BPSO ALGORITHM IMPLEMENTATION

α	N	SF	Best GF	CC (%)	α	N	SF	Best GF	CC (%)
0	10	17	0.1635	92.1	0.1	10	18	0.1624	93.4
	20	15	0.1442	94.6		20	14	0.1265	94.6
	30	11	0.1058	94.0		30	10	0.0925	94.3
	40	14	0.1346	94.6		40	13	0.1166	95.9
	50	11	0.1058	94.0		50	14	0.1256	95.6
0.2	10	18	0.1549	91.8	0.3	10	15	0.1208	93.4
	20	11	0.0934	95.6		20	15	0.1152	95.3
	30	14	0.1172	95.3		30	14	0.1026	94.9
	40	11	0.1035	90.1		40	11	0.0901	94.6
	50	13	0.1076	96.2		50	12	0.0997	93.7
0.4	10	15	0.1093	94.3	0.5	10	17	0.1133	93.4
	20	16	0.1138	94.6		20	14	0.0910	95.3
	30	12	0.0894	94.9		30	13	0.0925	94.0
	40	12	0.0919	94.3		40	15	0.0942	95.6
	50	11	0.0912	93.1		50	14	0.0862	96.2
0.6	10	15	0.0766	96.8	0.7	10	20	0.0842	96.2
	20	17	0.0900	95.9		20	18	0.0784	96.2
	30	12	0.0821	94.0		30	20	0.0820	96.5
	40	16	0.0843	96.2		40	17	0.0733	96.5
	50	16	0.0861	95.9		50	14	0.0691	95.9
0.8	10	25	0.0758	96.5	0.9	10	27	0.0572	96.5
	20	26	0.0727	97.1		20	31	0.0582	96.8
	30	19	0.0593	97.1		30	25	0.0496	97.1
	40	24	0.0663	97.5		40	28	0.0553	96.8
	50	24	0.0689	97.1		50	24	0.0515	96.8
1.0	10	47	0.0252	97.4	1.0	40	45	0.0284	97.1
	20	46	0.0284	97.1		50	54	0.0252	97.4
	30	50	0.0252	97.4					

The results show that when applying the BPSO algorithm for variable selection in the problem of classifying stable electrical systems, the single-objective function considering only the classification error, drives the algorithm to select a large number of variables. Therefore, applying the multi-objective function (4) for the variable selection problem in the classification of stable electrical systems is appropriate when using the BPSO algorithm.

VI. CONCLUSION

The paper proposed the application of the multi-objective function for the BPSO algorithm in selecting variables in the power system stability classification problem. The results from the experiment conducted on the IEEE-39 bus system demonstrated the success of the variable reduction technique. Reduction from 104 to 11 variables with the accuracy of up to 95.6% has significant economic implications, such as significantly reducing the cost of measuring devices to collect data. The proposed method opens up the direction of applying advanced computational techniques in selecting input parameters for the problem of classifying power system stability.

ACKNOWLEDGEMENT

This work belongs to the project T2022-45 funded by Ho Chi Minh City University of Technology and Education, Vietnam.

REFERENCES

[1] K. Teeparthi and D. M. V. Kumar, "Dynamic Power System Security Analysis Using a Hybrid PSO-APO Algorithm," *Engineering, Technology & Applied Science Research*, vol. 7, no. 6, pp. 2124–2131, Dec. 2017, https://doi.org/10.48084/etasr.1477.

[2] Y. V. Makarov, V. I. Reshetov, A. Stroev, and I. Voropai, "Blackout Prevention in the United States, Europe, and Russia," *Proceedings of the IEEE*, vol. 93, no. 11, pp. 1942–1955, Aug. 2005, https://doi.org/10.1109/JPROC.2005.857486.

[3] H. F. Khan, A. H. Hanif, and N. Anwar, "Rotor Angle and Voltage Stability Analysis with Fault Location Identification on the IEEE 9 Bus System," *Engineering, Technology & Applied Science Research*, vol. 10, no. 1, pp. 5259–5264, Feb. 2020, https://doi.org/10.48084/etasr.3175.

[4] D. V. Ngo, K. V. Pham, D. D. Le, K. H. Le, and K. V. Huynh, "Assessing Power System Stability Following Load Changes and Considering Uncertainty," *Engineering, Technology & Applied Science Research*, vol. 8, no. 2, pp. 2758–2763, Apr. 2018, https://doi.org/10.48084/etasr.1892.

[5] A. Sharifian and S. Sharifian, "A new power system transient stability assessment method based on Type-2 fuzzy neural network estimation," *International Journal of Electrical Power & Energy Systems*, vol. 64, pp. 71–87, Jan. 2015, https://doi.org/10.1016/j.ijepes.2014.07.007.

[6] S. M. Ashraf, A. Gupta, D. K. Choudhary, and S. Chakrabarti, "Voltage stability monitoring of power systems using reduced network and artificial neural network," *International Journal of Electrical Power & Energy Systems*, vol. 87, pp. 43–51, May 2017, https://doi.org/10.1016/j.ijepes.2016.11.008.

[7] 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," *Electric Power Systems Research*, vol. 148, pp. 264–272, Jul. 2017, https://doi.org/10.1016/j.epsr.2017.03.029.

[8] R. Zhang, Y. Xu, Z. Y. Dong, K. Meng, and Z. Xu, "Intelligent systems for power system dynamic security assessment: Review and classification," in *2011 4th International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT)*, Weihai, China, Jul. 2011, pp. 134–139, https://doi.org/10.1109/DRPT.2011.5993876.

[9] R. Zhang, Y. Xu, Z. Y. Dong, and D. J. Hill, "Feature selection for intelligent stability assessment of power systems," in *2012 IEEE Power and Energy Society General Meeting*, San Diego, CA, USA, Jul. 2012, https://doi.org/10.1109/PESGM.2012.6344780.

[10] S. Kalyani and K. S. Swarup, "Pattern analysis and classification for security evaluation in power networks," *International Journal of Electrical Power & Energy Systems*, vol. 44, no. 1, pp. 547–560, Jan. 2013, https://doi.org/10.1016/j.ijepes.2012.07.065.

[11] B. Xue, M. Zhang, and W. N. Browne, "New fitness functions in binary particle swarm optimisation for feature selection," in *2012 IEEE Congress on Evolutionary Computation*, Brisbane, QLD, Australia, Jun. 2012, https://doi.org/10.1109/CEC.2012.6256617.

[12] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95 - International Conference on Neural Networks*, Perth, WA, Australia, Aug. 1995, vol. 4, pp. 1942–1948. https://doi.org/10.1109/ICNN.1995.488968.

[13] J. Kennedy and R. C. Eberhart, "A discrete binary version of the particle swarm algorithm," in *Computational Cybernetics and Simulation 1997 IEEE International Conference on Systems, Man, and Cybernetics*, Orlando, FL, USA, Jul. 1997, vol. 5, pp. 4104–4108, https://doi.org/10.1109/ICSMC.1997.637339.

[14] J. D. Glover, M. S. Sarma, and T. Overbye, *Power Systems Analysis and Design*, 5th ed. 2012.

[15] A. Y. Abdelaziz and M. A. El-Dessouki, "International Journal of Intelligent Systems and Applications(IJISA)," *International Journal of Intelligent Systems and Applications(IJISA)*, vol. 5, no. 10, Sep. 2013, Art. no. 1.

[16] N. Amjady and S. A. Banihashemi, "Transient stability prediction of power systems by a new synchronism status index and hybrid classifier," *IET Generation, Transmission & Distribution*, vol. 4, no. 4, pp. 509–518, Apr. 2010, https://doi.org/10.1049/iet-gtd.2009.0255.

[17] A. Karami and S. Z. Esmaili, "Transient stability assessment of power systems described with detailed models using neural networks," *International Journal of Electrical Power & Energy Systems*, vol. 45, no. 1, pp. 279–292, Feb. 2013, https://doi.org/10.1016/j.ijepes.2012.08.071.