IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 12 July 2022, accepted 27 July 2022, date of publication 5 August 2022, date of current version 10 August 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3196769

RESEARCH ARTICLE

Fault Detection and Classification for Wide Area Backup Protection of Power Transmission Lines Using Weighted Extreme Learning Machine

ANI HARISH[1], (Senior Member, IEEE), A. PRINCE[1], (Senior Member, IEEE), AND M. V. JAYAN[2]
[1]Department of Electrical Engineering, Rajiv Gandhi Institute of Technology Kottayam, APJ Abdul Kalam Technological University, Thiruvananthapuram, Kerala 695016, India
[2]APJ Abdul Kalam Techological University, Thiruvananthapuram, Kerala 695016, India
Corresponding author: Ani Harish (ani.ramachandran@rit.ac.in)

ABSTRACT The changing landscape of power grids with distributed energy sources and power electronic devices has led to increasing relay maloperations. Wide area backup protection is necessary for the resolution of faults and for a reliable power grid. This paper presents detecting and classifying faults in transmission lines for wide-area backup protection using phasor measurement units (PMU) data. The faults are detected and classified using a Weighted Extreme Learning Machine (WELM) algorithm, which considers the variable distribution of data among the different classes using a weighted approach. The PMU signal data used were generated by the simulation of an IEEE 39 bus test system in the PowerWorld/OpenPDC/MATLAB environment. For classification, the input features data were derived using a wavelet transform-based ensemble feature extraction technique, and the WELM classifier was optimized using Particle Swarm Optimization (PSO). The PSO optimized WELM (PSO-WELM) model trained on PMU data detected faults with 100% accuracy and classified them into different types with an accuracy of 99.85%. It is validated that the PSO-WELM outperforms other known classifiers on performance comparison. The model also classified noisy data with a signal-to-noise ratio (SNR) as low as 10 dB and with an accuracy of 97%.

INDEX TERMS Data, fault, transmission lines, fault detection and classification, phasor measurement unit, extreme learning machine, machine learning.

I. INTRODUCTION

Present-day power grids are equipped with Phasor measurement units (PMUs) to collect GPS synchronized data. With the deployment of PMUs and other intelligent electronic devices on the power grid, a vast volume of measurement data is captured and stored. Valuable insights can be drawn from this burgeoning amount of data [1]. Data mining or machine learning algorithms help in extracting information from data. Due to the increasing adoption of distributed energy sources and power electronic devices on the power grid, many cases of maloperation of the relays are being reported. Most blackouts observed are due to the cascade impact of a primary protection system failure [2], [3]. In the wake of various blackouts that have happened worldwide, the need for a backup protection system is paramount for detecting disturbances and outages that can cascade and lead to a blackout [4], [5].

Transmission line protection is critical for power grid reliability. Transmission line protection solutions must include detecting transmission line faults quickly, establishing the type of fault, and pinpointing the problem. Machine learning algorithms yield data models that diagnose transmission line faults by mining power system measurement data [6]. References [7] and [8] give a comprehensive literature review of the different methods that are used for locating and classifying transmission line faults. The faults detection and classification (FD&C) approaches for transmission lines can be categorized based on mathematical system models, knowledge-based models, and data-driven models [6]. Developing a reliable mathematical system model has become complex with the proliferation of stochastic renewable energy sources in the power grids [6]. Intelligent methods that derive information from the measurement data are most suitable for

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

developing protection solutions for the current power grid. Data mining applications can be essential components of power systems' wide area backup protection (WABP).

Different feature extraction techniques help to derive the necessary patterns from measurement data for building machine learning models and extract pertinent information from the power system signal data for fault detection and classification. Signal processing techniques, Wavelet transform, Wavelet multivariate analysis, and Wavelet entropy analysis are the most applied feature extraction methods [9]–[12] for FD&C in transmission lines. In [13] and [14], faults are classified using the energy values of the wavelet approximation coefficients. Reference [15] proposes an ensemble feature extraction method using wavelet transform. Fast Fourier Transform (FFT), S-transform, and Hilbert Huang transform are other data transformation methods used for feature extraction [16]–[18]. The machine learning algorithms like Artificial Neural Networks (ANN), Support Vector Machines (SVM), K Nearest Neighbor (KNN), Decision tree (DT), and Bagging and Boosting Ensemble Tree (BTEC) are widely used for faults classification and localization [11], [19]–[23]. Reference [24], [25], and [26] describe transmission line faults classification methods with integrated feature extraction. In [24], the authors propose a fault classification technique using Hidden Markov Model. [25] presents an unsupervised learning convolutional sparse encoder method for fault classification. Reference [26] offers a combined fault categorization and localization method using Extreme Learning Machines (ELM). The classification models of [24], [25], and [26] are trained on data sets generated by faults simulated on a single transmission line.

A. RELATED WORK

Wide-area measurements offer a comprehensive view of the system and can be used to create FD&C models for wide-area backup protection. PMUs provide GPS synchronized wide-area measurements and are used for developing FD&C models. Reference [27] describes detecting and classifying events on a power grid using PMU data. This research focuses only on fault detection and does not consider categorizing faults into different types. In [16], the researchers propose a fault recognition and categorizing method using PMU data. Park and Fast Fourier Transformation method extracts features, and SVM classifies the faults. Researchers in [16] have not quantified the time to train and test the model. However, the training time complexity of SVM with large datasets is high, and they require sizeable amounts of memory and processing power. Furthermore, SVM classifiers have some limitations when dealing with multiclass classification problems and imbalanced datasets [28], [29].

In [30], feature extraction is done through phaselet transform and faults classification by the Gaussian Naive Bayes (GNB) algorithm. Data used is PMU measurements from two ends of a transmission line. Researchers in [13] propose a Discrete Wavelet transform-based scheme for transmission line protection. Features extracted are the energy values of the wavelet decomposition. In [30] and [13], the data sets considered are from PMUs placed on both transmission line ends. PMU deployment on a transmission line necessitates installing the sensing devices (current and potential transformers) and the communication infrastructure with the measurement unit. The capital cost of installing PMUs on a power system's transmission lines will be very high. PMUs are typically mounted on buses to save costs. A substantial body of research has suggested ways to determine the ideal number of buses for PMU placements on a power system [31].

The majority of the machine learning methods for developing FD&C models for detecting, classifying, and localizing faults on transmission lines are developed for transmission lines with specific topology or with data from a single end or both ends of the transmission line [9]–[12], [19]–[22], [23]. References, [13], [16], [27], [30] are some of the works that present FD&C models developed using PMU data. However, there is still a need for computationally efficient methods to develop FD&C models for WABP, considering the availability of fewer PMU measurements due to a reduced selected number of PMU placements on the power grid and noise present in the measurements.

The data distribution among different classes in the input data set used in training a classifier model plays a vital role in influencing the performance of the classification model [32]. A data set with more than one class, which has more samples in one class than the average number of samples across the different classes, is an imbalanced dataset. The class with more records is the majority and the other classes are the minority classes. A classifier model with an imbalanced input dataset can have high accuracy if it classifies the majority class correctly, although the predictions for the minority classes may be poor [33]. Most of the time, a power system operates under no-fault conditions, so the number of measurements in the No-fault class will be higher than the number of measurements in the fault classes, leading to an unbalanced data distribution for the fault classes in the PMU signals data set. It is essential to consider the imbalanced distribution of data across classes in the PMU data set while developing FD&C data-driven models. To the best of our knowledge, there is very little study of this aspect of imbalanced data sets in the literature while developing FD&C models for transmission lines using PMU data.

B. CONTRIBUTION

This paper presents FD&C models for wide area backup protection of transmission lines using PMU data and Extreme Learning Machines. Extreme Learning Machines (ELM) are Neural Networks with a single layer. They are computationally swift compared to Artificial Neural Networks with Multi-Layer Perceptrons (MLPs). Due to the presence of a single layer, they do not have the local minimum problem of Backpropagation MLP Networks [34]. ELMs are popular and find applications in many fields [35]. Optimizing the input weights and biases can improve the performance of the

ELM algorithm. Particle Swarm Optimization (PSO) can be used to optimize the value of input weights of an Extreme Learning Machine to improve its performance [36], [37].

The main contributions of this research work are

- This paper presents an optimized WELM (PSO-WELM) method for developing fault diagnosis models for WABP of power systems using PMU data. The PSO-WELM approach considers the uneven distribution of PMU data among classes while developing the models. The ability of classifier models to generalize and learn while categorizing unbalanced input data sets is enhanced when the data distribution among the classes of the input data set is considered while developing the models.
- The performance of the PSO-WELM models is evaluated with complete, sparse, and noisy measurement data.
- An ensemble feature extraction method is used to extract the most relevant features from PMU data and to improve the performance of FD&C models.

C. STRUCTURE OF THE PAPER

The remaining part of this paper is organized as follows. Section ll describes the methodology for developing power system FD&C machine learning models, ELM, WELM, and PSO theory. Section III gives the simulation setup for data generation and describes the ensemble features extraction method and faults detection and classification with the proposed PSO-WELM method. Section IV has the results and discussion, and Section V has the conclusion.

II. METHODOLOGY, EXTREME LEARNING MACHINE, WEIGHTED EXTREME LEARNING MACHINE AND PARTICLE SWARM OPTIMIZATION (PSO)

A. METHODOLOGY FOR DEVELOPING POWER SYSTEM FD&C MACHINE LEARNING MODELS

Fig 1 illustrates the generic methodology for developing data-driven fault diagnosis system models for power systems. FD&C models can be built using data from real power grids or simulations of test systems created for research and testing. Most data for constructing machine learning models for power grids is obtained through simulation because it is possible to mimic all fault scenarios whose frequency of occurrence is less in real-world measurements. Data pre-processing follows data collecting and includes checking for data accuracy, cleaning, labeling, and normalizing. The data is transformed, and features that give helpful information about it are extracted. Pattern matching machine learning-based models are then used to detect and categorize faults into distinct types using the features data set.

B. EXTREME LEARNING MACHINES (ELM)

ELMs are feedforward networks with a single layer [34], [35], [38] with output as in (1).

$$f(x) = \sum_{j=1}^{N} B_j\, hi_j(x) = Hi\, B_j \tag{1}$$

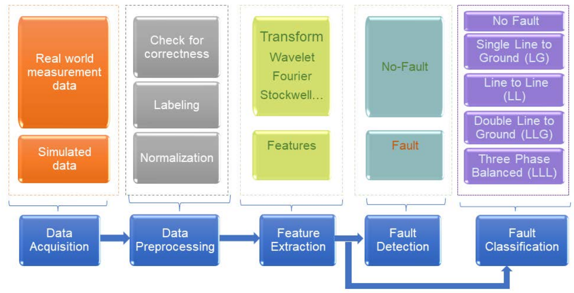

FIGURE 1. Methodology for developing power system FD&C models.

where $B = [B_1, B_2, \cdots, B_N]$ are the weights between the hidden layer and the output layer, and $Hi = [hi_1(x), hi_2(x), \cdots, hi_N(x)]^T$ are the output of the hidden layer nodes to the output layer nodes. hi(x) uses a non-linear mapping function to map the inputs to a feature space as in (2), and N is the number of input samples, and L is the number of hidden neurons. The mapping function commonly used is the sigmoid function, as in (3)

$$hi_j(x) = F(s_i, t_i, x),\ s_i \in R^d,\ t_i \in R \tag{2}$$

$$F(s, t, x) = \frac{1}{-exp(sx + t)} \tag{3}$$

The ELM learning minimizes the squared error between the inputs and the targets. Moore Penrose generalized inverse of H_i gives the output weights B as in (4).

$$B = \hat{H}iT \tag{4}$$

where T is the target output matrix.

When $N < L$ then,

$$B = Hi^T \left(\left(\frac{1}{C} \right) + HiHi^T \right)^{-1} T \tag{5}$$

When $N > L$

$$B = \left(\left(\frac{1}{C} \right) + HiHi^T \right) Hi^T T \tag{6}$$

For better generalization, a positive value C is added to $HiHi^T$. C is a Regularization parameter that makes the solution more robust.

C. WEIGHTED EXTREME LEARNING MACHINE (WELM)

In a multiclass classification problem, a data set has a skewed structure when a class has more than the average number of records across all classes. The majority class has more records than the minority class. In [39], the authors propose a Weighted Extreme Learning Machine algorithm for imbalanced data classification. A weighting matrix, W, is defined for every input sample x_i to address the skewed data problem. The weight Wii will be more significant for a minority class than the majority class. The output B of the hidden neurons is When $N < L$ then,

$$B = Hi^T \left(\left(\frac{1}{C} \right) + WHiHi^T \right)^{-1} T \tag{7}$$

When $N > L$

$$B = \left(\left(\frac{1}{C}\right) + Hi^T WHi\right)^{-1} Hi^T WT \quad (8)$$

The weights W_{ii} are approximated in two different ways. In the first weighting method, Wii is as in (9)

$$W_{ii} = \frac{1}{\#n_c} \quad (9)$$

Second weighting method as in (10),

$$\begin{aligned} W_{ii} &= \frac{0.618}{\#n_c} \qquad if\ n_c > Avg(n_c) \\ &= \frac{1}{\#n_c} \qquad if\ n_c <= Avg(n_c) \end{aligned} \quad (10)$$

where n_c is the number of instances in a class.

D. PARTICLE SWARM OPTIMISATION (PSO)

In PSO, particles consisting of the swarm move with a velocity in the search space to find the best position. Each particle in the swarm adjusts its movement according to its experience and the experience of the other particles [40], [41].

The velocity vector of the ith particle is expressed as $\hat{\upsilon_i} = [\hat{\upsilon_{i,1}}, \hat{\upsilon_{i,2}} \ldots \hat{\upsilon_{i,n}} \ldots \hat{\upsilon_{i,D}}]$ and the position of the ith particle is $\hat{\rho_i} = [\hat{\rho_{i,1}}, \hat{\rho_{i,2}} \ldots \hat{\rho_{i,n}} \ldots \hat{\rho_{i,D}}]$. $\hat{\rho_{best,i}}$ is the best position of the ith particle, and $\hat{g_{best}}$ is the best position of all the particles.

The position and velocity of each particle evolve as per the formula in (11 and 12).

$$\begin{aligned} \hat{\upsilon}_i(n+1) &= \hat{\omega}\hat{\upsilon}_i(n) + c_{one}r_{one}(\hat{\rho_{best,i}}(n) - \hat{\rho}_i(n)) \\ &\quad + c_{two}r_{two}\hat{g_{best}}(n) - \hat{\rho}_i(n)) \end{aligned} \quad (11)$$

where $1 <= i <= n_p$, $1 <= n <= D$ and,

$$\hat{\rho}_i(n+1) = \hat{\rho}_i(n) + \hat{\upsilon}_i(n+1) \quad (12)$$

Here, n is the number of iterations, n_p the number of particles, D is the maximum number of iterations, $\hat{\omega}$ is the inertial weight, c_{one} and c_{two} are acceleration constants, r_{one} and r_{two} are random numbers with a uniform distribution in the range [0,1].

The performance of PSO is greatly influenced by the choice of acceleration constants c_{one} and c_{two} and the inertial weight $\hat{\omega}$. The inertial weight $\hat{\omega}$ varies within the range of ω_{min} and ω_{max}. The inertial weight ω is calculated as in (13).

$$\hat{\omega} = \omega_{max} - \left\{\left\{\frac{\omega_{max} - \omega_{min}}{D}\right\} * n\right\} \quad (13)$$

III. FAULT DETECTION AND CLASSIFICATION WITH PSO-WELM

A. SIMULATION SETUP FOR DATA ACQUISITION

For Data Acquisition, an IEEE 39 bus system was simulated in the PowerWorld Simulator [42]. The system parameters for simulating IEEE 39 bus test system are from [43].

Faults were simulated in the IEEE 39 bus test system using the Transient stability module of the PowerWorld Simulator. The Transient stability module of the PowerWorld Simulator has inbuilt predefined fault type elements. For each fault case, the simulation was run for 5 sec, and faults were inserted and cleared after one to two cycles. The different fault case data were generated by varying the location of the fault, the type of fault, and the resistance. Table 1 lists the various faults simulation parameters.

TABLE 1. Faults simulation parameters.

Parameter	Values
Fault type	No-fault, LG, LL, LLG, LLL
Fault Resistance	1, 5, 10, 15, 20, 25, 35, 40, 50 ohms
Fault location	10, 20, 30, 40, 50, 60, 70, 80, 90% of line length

On the simulation run, the Dynamic simulator of the PowerWorld Simulator generated IEEE C37.118 compliant messages, which were collected by OpenPDC, an open-source Phasor Data Concentrator (PDC). In OpenPDC, the messages are stored in the Historian/SQLserver database. The .csv files generated by the OpenPDC Historian are imported into the MATLAB workspace. The data is then checked for inconsistencies, cleaned, and labeled. The simulation setup with the PowerWorld simulator, OpenPDC, and MATLAB is illustrated in Fig. 2

The data set generated has 24,654 samples of voltage, voltage-angle, current, current-angle, and frequency signals from the 39 buses. The generated dataset was split using 70:30 proportion into training data set with 15317 samples and a testing data set with 7134 samples.

TABLE 2. Data Distribution for Fault Detection.

Data set	NF (1)	Fault (2)
Training	3868	11449
Testing	2229	4905

B. FEATURE EXTRACTION

Fig.3 illustrates the feature extraction and FD&C process flow. Features are meaningful information about the input data. Many research works on fault analysis of power system data have proposed Wavelet transformation as the most efficient feature extraction technique for power signal data [11]–[14]. The Discrete Wavelet Transform (DWT) was used for transforming the signals data. The Daubechies Db4 mother wavelet was used for decomposing the signals into five levels of wavelet coefficients. The features data set is an ensemble of the following

- Feature-set 1 comprises the statistical features, Standard deviation, Range, Root mean square value (RMS), and Crest Factor of the signals data.
- Feature-set 2 comprises the Entropy values of the wavelet coefficients of the signal data set calculated using (14). Shannon Entropy quantifies the amount of

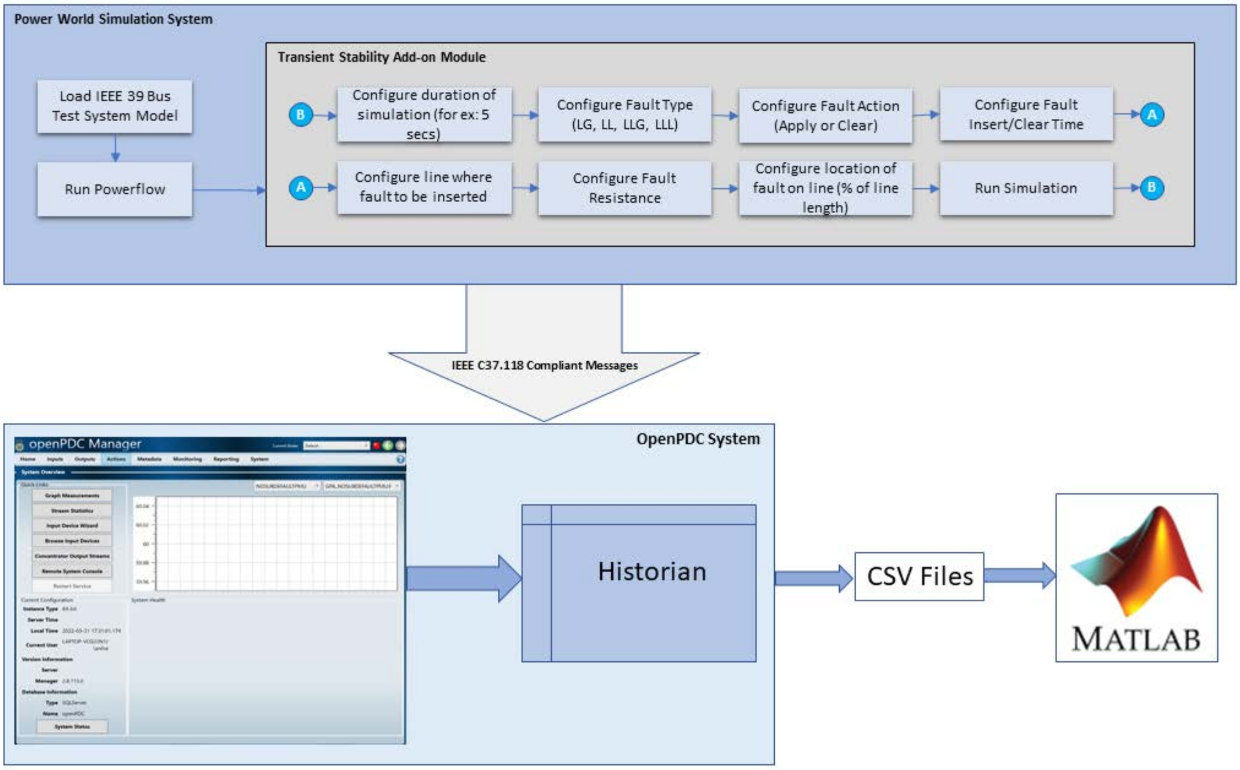

FIGURE 2. Simulation setup for dataset generation.

TABLE 3. Data Distribution for different Fault Classes.

Data set	NF (1)	LG (2)	LL (3)	LLG (4)	LLL (5)
Training	3868	2694	2694	2694	3367
Testing	2229	1154	1154	1154	1443

information in a variable. The Shannon Entropy of the wavelet coefficients at level j is calculated as per (14)

$$WSE_j = \sum_{i=1}^{N} p_{j,k} \log p_{j,k} \tag{14}$$

where N is the number of coefficients in the jth level and $p_{j,k}$ normalized squares of the wavelet coefficients at the jth level.

- Feature-set 3 comprises the energy (L2 norm) of each decomposition. The wavelet energy value is calculated as in (15).

$$WaveletEnergy = \||ca||^2 + \||cd||^2 \tag{15}$$

where ca is the approximation coefficient, and cd is the detail coefficient of the wavelet decomposition. The plots of the Shannon entropy and the energy values of the wavelet coefficients of the three phase currents and voltages for the different faults are illustrated in Fig. 4.

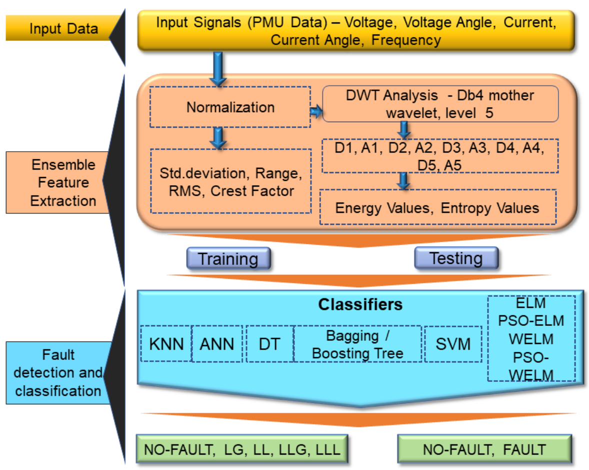

FIGURE 3. Ensemble Feature Extraction, Fault Detection, and Classification.

C. PSO ALGORITHM FOR WELM WEIGHTS OPTIMIZATION

The flowchart of the PSO-WELM algorithm is as in Fig.5. The optimization of WELM network weights with PSO is described in the following steps.

- Initialize the activation function f(x) and the number of hidden layer neurons. Initialize the swarm size (n_p). Input the training dataset and randomly initialize each particle in the swarm with an input weight for the

ELM network within the lower bound value for weights *LB* and upper bound values for weights *UB*. Initialize the PSO parameters. The parameters for the PSO chosen for FD&C are as in Table 4.

- Calculate the corresponding output weights (B) for each particle using equations (7) or (8).
- Calculate the output fitness value, F_{value} as per (16)

$$F_{value} = (Inputs * B') * pinv(B') \tag{16}$$

- Calculate the fitness of each particle as per (17)

$$fitness_i = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}\left(F_{value} - Inputs\right)^2}, \tag{17}$$

where n is the number of samples in the training dataset.

- Set the best position of each particle $\rho_{\hat{best},i}$ with its initial position and the global best $g_{\hat{best}}$ as the particle with the minimum fitness value among all the search trajectories.
- Update the velocity and position of each particle using (11) and (12).
- Repeat the optimization process till the maximum number of iterations,
- The best global best $g_{\hat{best}}$ gives the best input weights for the WELM network.
- Train the WELM classifier model with the optimal input weights.
- Test the WELM model with the testing dataset.

TABLE 4. PSO parameters for WELM weights optimization.

Description	Parameter	Value
Size of population	n_p	50
Maximum no. of iterations	n_{max}	500
Acceleration constants	c_{one}, c_{two}	2
Initial inertial weight	ω_{max}	0.9
Final inertial weight	ω_{min}	0.2
Lower bound value for weights	LB	-1
Upper bound value for weights	UB	1

D. FAULT DETECTION AND CLASSIFICATION

Fault detection is a binary classification problem with the two classes being No-fault (class 1) and Fault (class 2). The weights of the WELM fault detection model are optimized with PSO. The PSO-WELM fault detection model was trained on the training data set and the performance of the classifier model was validated with the testing data set in Table 2.

The fault classification is a multiclass classification problem. The different classes in the data set are No-fault (NF), Line to Ground fault (LG), Line to Line fault (LL), Line to Line to Ground fault (LLG), and three phase balanced fault (LLL). The PSO-WELM fault classification model was trained and tested with the data sets having the data distribution among different classes as in Table 3. As can be observed from Table 3, there is an unequal data distribution among the different classes, and so a weighted approach is adopted for initializing the weights of the ELM network.

E. SPARSE DATA

Due to economic constraints, it may not be feasible to install PMUs on all the power system buses [31]. We used the PSAT toolbox for MATLAB for PMU placement [44]. Different cases considered are

- Case 1: PMUs on 16 buses. Data collected are signals from the 2, 6, 8, 10, 12, 14, 16, 18, 20, 23, 27, 33, 35, 37, 38, and 39 buses. Depth First search optimization algorithm used to identify the optimal number of buses for the IEEE 39 bus test system.
- Case 2: PMUs on 14 buses. Data collected are signals from the 4, 8, 16, 28, 31, 32, 33, 34, 35, 36, 37, 38, and 39 buses.
- Case 3: PMUs on 11 buses. Data collected are signals from the 2, 6, 9, 10, 14, 16, 20, 23, 26, 37, and 38 buses. Graph-theoretic procedure optimized the IEEE 39 bus system to select the 11 buses for PMU placement.
- Case 4: PMUs on 9 buses. Data collected are signals from the generator buses 31, 32, 33, 34, 35, 36, 37, 38, and 39.

F. DATA WITH NOISE

To emulate the noise levels in real field measurements, white Gaussian noise of SNR 10 dB to 60 dB were added to the different training and testing datasets. Noisy data sets were created with complete as well as sparse measurement data sets, and FD&C models were developed for all the noisy data sets. The presence of noise in real world PMU measurements was studied in [45] and [46]. It is shown in [45] that the noise distribution in field PMU data is Gaussian with a Signal to noise ratio (SNR) of 45dB and more.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

The PSO-WELM fault detection model detects faults with an accuracy of 100%. The confusion matrix for fault detection, categorizing data into NF (1) and Fault (2) classes is as in Fig. 7.

The PSO-WELM classifier groups the fault classes with an accuracy of 99.85%. The confusion matrix for fault classification into NF (1), LG (2), LL (3), LLG (4), and LLL (5) classes is as in Figure 8.

A. PERFORMANCE COMPARISON OF PSO-WELM WITH OTHER CLASSIFIERS

PSO-WELM's performance is compared to that of ELM, PSO-ELM, WELM, DT, KNN, GNB, SVM, and BTEC classifiers. When classifying data sets with imbalanced class distributions, accuracy alone will not suffice as a performance metric. The different metrics considered for comparison of model performances are Precision, Recall, F1 score, Training, and Testing time in addition to Accuracy. The equations for computing the performance measures are given in (18) - (21).

$$Accuracy(A_{cc}) = \frac{\Psi_t + \hat{\Psi}_t}{\Psi_t + \hat{\Psi}_t + \Psi_n + \hat{\Psi}_n} \tag{18}$$

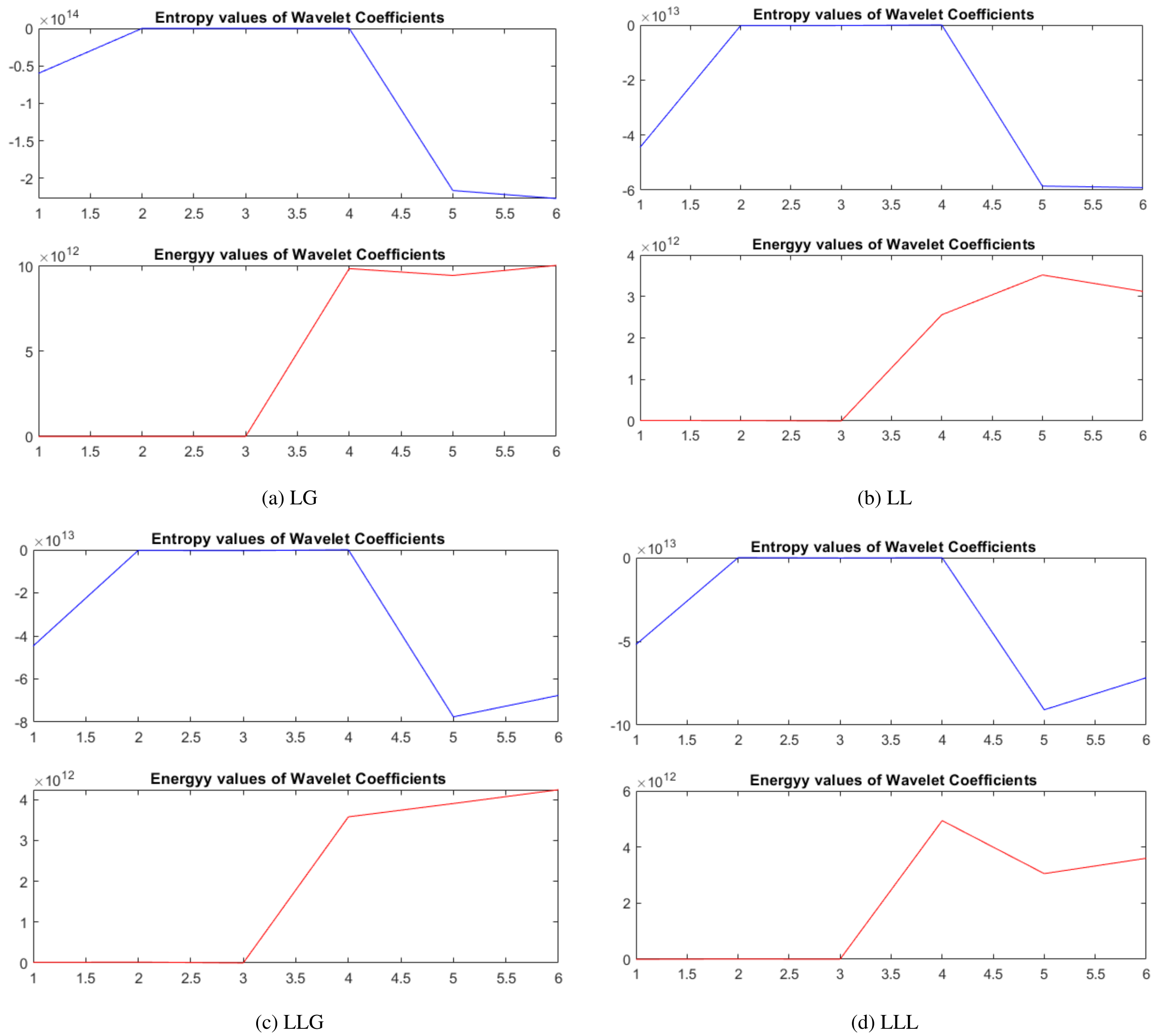

FIGURE 4. Plot for the energy and entropy values of wavelet coefficients of current and voltage signals for different faults. Energy values vary for the third, fourth and the fifth levels of detail coefficients for different fault types. Entropy value variations can be observed for the approximation coefficient (value 1 in X axis) and the fourth and fifth levels of wavelet coefficients.

$$Precision(P_{rcn}) = \frac{\Psi_t}{\Psi_t + \hat{\Psi}_t} \tag{19}$$

$$Recall(R_{cl}) = \frac{\Psi_t}{\Psi_t + \hat{\Psi}_n} \tag{20}$$

where Ψ_t is the number correctly classified as the true class, $\hat{\Psi}_t$ is the number incorrectly classified as the true class, Ψ_n is the number correctly classified as the negative class, and $\hat{\Psi}_n$ is the number incorrectly classified as the negative class.

$$F1score = 2 * \frac{P_{rcn} * R_{cl}}{P_{rcn} + R_{cl}} \tag{21}$$

Table 5 details the values of the performance metrics of the different classifiers with the complete system-wide measurement data and sparse measurement data. The models were trained and tested on a PC with Intel(R) Core(TM) i7-9750H CPU @ 2.60GHz, 64-bit operating system, x64-based processor and 32 GB RAM.

The performance of the FD&C models with sparse measurement data sets is as in Figure 6. The accuracy of classification is above 96% for all the models at 40dB noise level.

The generalization capability of a machine learning model is how well the model classifies unseen data. The PSO-WELM model was trained and tested on the training and testing dataset as in Table 3, and the model classified the hitherto unknown testing dataset with an accuracy above 99.85%. The generalization capability of the PSO-WELM classifier model is substantial, as demonstrated by the testing accuracy. The WELM network weights are optimized using

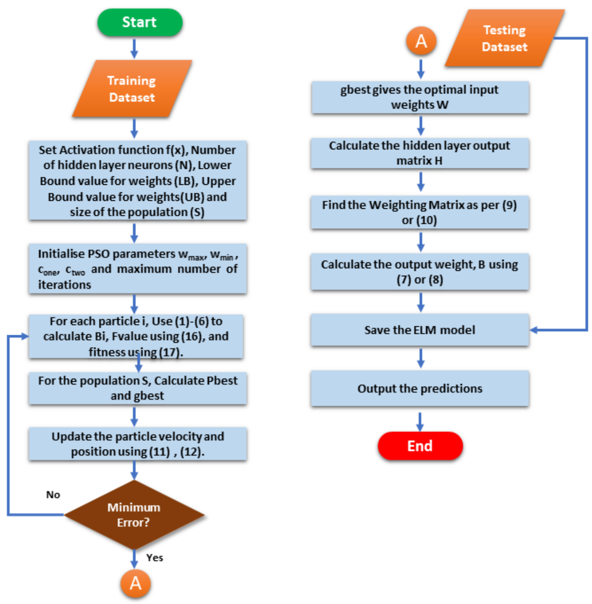

FIGURE 5. Flowchart of PSO-WELM.

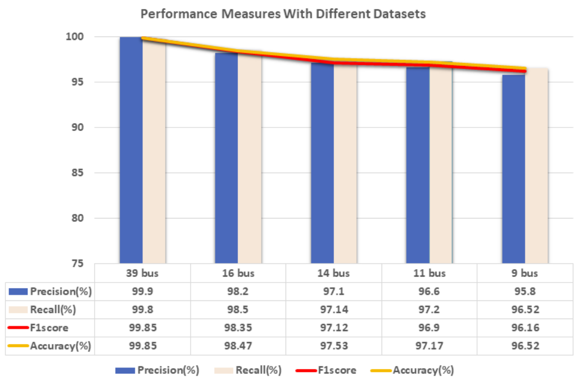

FIGURE 6. Accuracy, Precision, Recall, and F1-Score of PSO-WELM Classifier with different data sets.

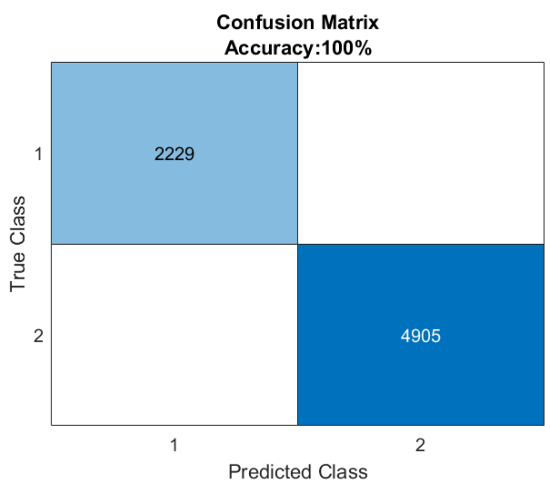

FIGURE 7. Confusion Matrix for fault detection.

Particle Swarm Optimization, and the usage of optimal values of weights during training improved the generalization capability of the WELM network. The improvement in the classification accuracy of PSO-WELM from WELM in classifying the testing dataset can be observed in Table 5. Due to the weighted approach, WELM and PSO-WELM classifiers outperform ELM and PSO-ELM classifiers.

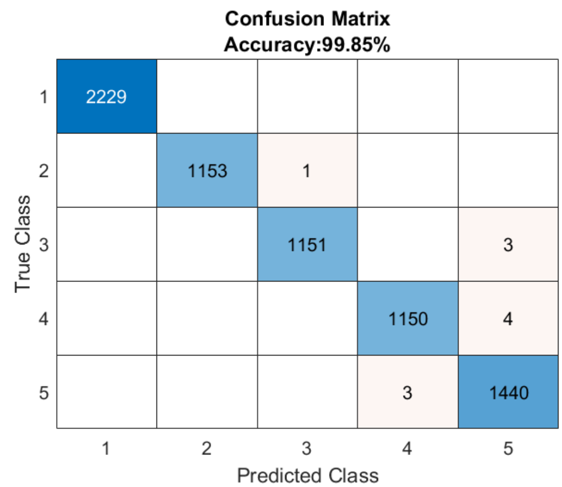

FIGURE 8. Confusion Matrix for fault classification with PSO-WELM.

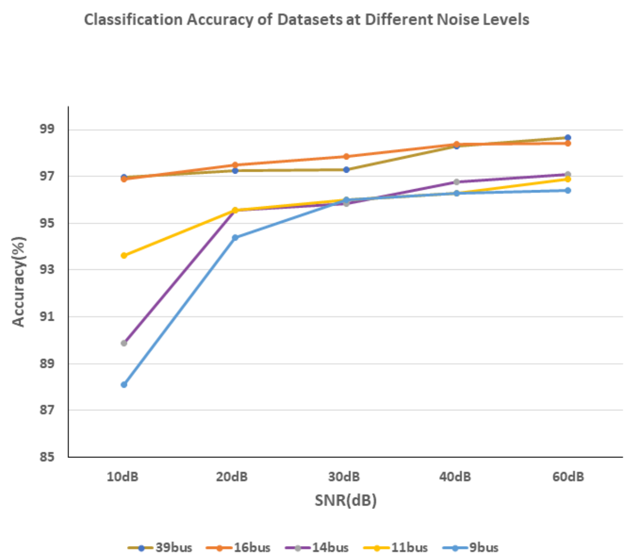

FIGURE 9. Accuracy of PSO-WELM classifier with different data sets at noise levels SNR 10dB-60dB.

SVM and KNN classifier models classify the PMU data from 39 buses with almost 95% accuracy, but training these models takes a long time. The time to train ELM-based models is significantly less than the DT, KNN, SVM, GNB, and BTEC driven models.

The overall performance of models trained with signals data from 16 buses is comparable with the models trained with the 39 bus signals dataset. The buses selected for placing the 16 PMUs, make the system fully observable, reducing redundancy of measurements and thus the efficacy of the 16 bus data set and the Depth First Search algorithm for identifying the optimal number of PMU placements for full

TABLE 5. Comparison of performance of classifiers with different measurement data sets.

Measurement data from	Method	Precision	Recall	F1score	Accuracy(%)	Training time(s)	Testing time(s)
39 buses	DT	0.86	0.81	0.84	84.68	3.72	0.07
	KNN	0.94	0.95	0.95	94.88	68.28	8.92
	SVM	0.96	0.94	0.95	95.04	11.95	0.24
	BTEC-Bagging	0.93	0.95	0.94	93.99	317.30	0.27
	GNB	0.60	0.60	0.60	51.77	11.21	0.31
	ELM	0.48	0.56	0.51	47.74	3.67	0.66
	PSO-ELM	0.48	0.57	0.52	48.07	4.11	0.64
	WELM	0.99	0.99	0.99	99.60	2.64	0.20
	PSO-WELM	**0.99**	**0.99**	**0.99**	**99.85**	**2.79**	**0.14**
16 buses	DT	0.80	0.83	0.84	82.75	3.32	0.03
	KNN	0.90	0.92	0.92	91.04	24.49	4.34
	SVM	0.89	0.93	0.91	91.28	14.76	0.47
	BTEC-Bagging	0.84	0.89	0.86	86.45	126.66	0.94
	GNB	0.55	0.54	0.54	47.80	5.23	0.09
	ELM	0.85	0.86	0.86	87.26	2.81	0.28
	PSO-ELM	0.65	0.69	0.67	56.74	2.72	0.20
	WELM	0.96	0.97	0.96	96.54	2.75	0.22
	PSO-WELM	**0.98**	**0.99**	**0.98**	**98.47**	**2.70**	**0.12**
14 buses	DT	0.79	0.81	0.80	82.48	11.73	0.16
	KNN	0.88	0.90	0.89	90.08	146.35	3.77
	SVM	0.87	0.88	0.88	88.94	344.45	1.21
	BTEC-Bagging	0.96	0.97	0.96	96.93	13.85	0.51
	GNB	0.50	0.48	0.49	43.30	3.22	0.10
	ELM	0.64	0.61	0.62	54.64	1.12	0.10
	PSO-ELM	0.80	0.81	0.81	83.17	0.58	0.05
	WELM	0.97	0.97	0.97	97.20	2.78	0.19
	PSO-WELM	**0.97**	**0.97**	**0.97**	**97.53**	**2.62**	**0.12**
11 buses	DT	0.76	0.76	0.76	79.00	10.26	0.03
	KNN	0.90	0.91	0.91	91.60	106.20	2.83
	SVM	0.82	0.84	0.83	84.59	113.36	0.13
	BTEC-Bagging	0.93	0.95	0.94	94.41	9.20	0.13
	GNB	0.34	0.26	0.29	30.12	2.72	0.07
	ELM	0.81	0.81	0.81	82.00	2.58	0.14
	PSO-ELM	0.77	0.78	0.78	80.80	2.52	0.39
	WELM	0.95	0.96	0.95	94.80	2.62	0.11
	PSO-WELM	**0.97**	**0.97**	**0.97**	**97.17**	**2.61**	**0.11**
9 buses	DT	0.75	0.77	0.76	78.32	8.19	0.25
	KNN	0.83	0.85	0.84	86.12	82.61	2.89
	SVM	0.80	0.82	0.81	83.19	44.80	0.42
	BTEC-Bagging	0.93	0.93	0.93	93.50	18.44	0.22
	GNB	0.48	0.47	0.47	41.00	2.24	0.08
	ELM	0.83	0.84	0.83	85.53	2.40	0.16
	PSO-ELM	0.82	0.83	0.83	84.74	2.67	0.16
	WELM	0.94	0.94	0.94	94.91	2.14	0.13
	PSO-WELM	**0.96**	**0.96**	**0.96**	**96.52**	**2.33**	**0.11**

observability of the system. As the dimensionality of input data sets reduces the computational requirements for training the machine learning models also decrease and the learning ability improves.

B. COMPARISON WITH OTHER REPORTED FEATURE EXTRACTION AND CLASSIFIER METHODS

The performance of the proposed classifier method (PSO-WELM) in categorizing the ensemble feature data set is compared with the classifier methods reported in the literature [9], [11], and [14] in classifying different feature data sets. In [9], the researchers propose feature extraction using wavelet entropy and Probabilistic Neural Network (PNN) for classification. In [11], the authors use the entropy values of the wavelet coefficients of the input data set as features and SVM as the classifier. In [14], the researchers propose the energy value of the wavelet decomposition coefficients as features and ANN for classification. Table 6 gives the different performance metrics of the PSO-WELM classifier in classifying ensemble features data set and the classifiers in [9], [11], and [14]. The SVM model classifying wavelet entropy features has good accuracy, but the time for training the SVM model is high. ANN and ELM models

TABLE 6. Comparison with other reported works.

Classifier	PSO-WELM	PNN	SVM	ANN
Features	Ensemble features	Wavelet Entropy	Wavelet Entropy	Wavelet Energy
Reference	Proposed method	[9]	[11]	[14]
Accuracy	99.85	88.76	95.49	92.47
Precision	99.90	86.18	94.46	91.24
Recall	99.80	90.86	95.91	91.34
Training Time	2.79	159.39	12.60	9.38
Testing Time	0.14	77.81	0.41	0.02

take comparatively less training time and the PSO-WELM model has the least training time and the best classification accuracy.

V. CONCLUSION

A PSO-WELM classifier model using an ensemble feature input data set derived from PMU data is designed and developed to detect and classify transmission line faults. The PSO-WELM-driven fault detection model detects faults with 100% accuracy, while the fault classification model has a 99.85% accuracy. When the data distribution among the different classes is taken into account, the detection and classifier models' generalization and learning abilities improve. PSO-WELM employs a weighted approach to increase the significance of minority classes, thereby improving model classification accuracy. Furthermore, this is demonstrated by contrasting performance metrics of standard ELM models with Weighted ELM models. The ELM models had high training accuracy but failed to accurately classify the test data set due to over-fitting and skewed unequal data distribution among the classes.

The time required for training the classification models with PSO-WELM is significantly less than DT, KNN, GNB, SVM, and BTEC Classifiers. The models perform in an acceptable range of classification accuracy of more than 96% with sparse and noisy data. Given the efficacy of the PSO-WELM method even in highly adverse scenarios, as evidenced in the analysis of the experimental results, PSO-WELM models will help reduce the overall capital cost of WABP systems. Mainly enabling WABP systems to reduce or optimize PMU placements in power systems and fast and accurate faults detection and classification protect the system from the cascade impact of maloperation of relays.

The PSO-WELM FD&C models have been trained and tested with data sets generated using an industry-accepted power system simulator. Noisy data sets were used to train the model to emulate the presence of noise in actual field measurement data. Although industry-standard simulators can offer characteristics and qualities of measurement data that are strikingly comparable to those in the real world, all of the characteristics of data variability resulting from dynamics in the actual world cannot be entirely reflected in simulated data sets. Lack of availability of public data sets with field measurements is a challenge when developing machine learning models for power systems FD&C. Future research directions can be to utilize the method for developing fault localization models using PMU data and using the PSO-WELM method to train models for backup protection of low voltage distribution systems using micro PMU (μ-PMU) data. Furthermore, it will be interesting to investigate other optimization methods to optimize WELM weights for training FD&C models.

REFERENCES

[1] A. Srivastava, K. G. Ravikumar, and G. Zweigle, "Wide-area monitoring and control using the real time digital simulator and a synchrophasor vector processor," *Eur. Trans. Electr. Power*, vol. 21, no. 4, pp. 1521–1530, May 2011, doi: 10.1002/etep.548.

[2] M. K. Neyestanaki and A. M. Ranjbar, "An adaptive PMU-based wide area backup protection scheme for power transmission lines," *IEEE Trans. Smart Grid*, vol. 6, no. 3, pp. 1550–1559, May 2015.

[3] S. A. Lavand and S. A. Soman, "Predictive analytic to supervise zone 1 of distance relay using synchrophasors," *IEEE Trans. Power Del.*, vol. 31, no. 4, pp. 1844–1854, Aug. 2016.

[4] J. J. Chavez, N. V. Kumar, S. Azizi, J. L. Guardado, J. Rueda, P. Palensky, V. Terzija, and M. Popov, "PMU-voltage drop based fault locator for transmission backup protection," *Electr. Power Syst. Res.*, vol. 196, Jul. 2021, Art. no. 107188, doi: 10.1016/j.epsr.2021.107188.

[5] A. G. Phadke, P. Wall, L. Ding, and V. Terzija, "Improving the performance of power system protection using wide area monitoring systems," *J. Mod. Power Syst. Clean Energy*, vol. 4, no. 3, pp. 319–331, Jul. 2016.

[6] R.-A. Tirnovan and M. Cristea, "Advanced techniques for fault detection and classification in electrical power transmission systems: An overview," in *Proc. 8th Int. Conf. Modern Power Syst. (MPS)*, May 2019, pp. 1–10.

[7] S. Brahma, R. Kavasseri, H. Cao, N. R. Chaudhuri, T. Alexopoulos, and Y. Cui, "Real-time identification of dynamic events in power systems using PMU data, and potential applications—Models, promises, and challenges," *IEEE Trans. Power Del.*, vol. 32, no. 1, pp. 294–301, Feb. 2017.

[8] A. Prasad, J. Belwin Edward, and K. Ravi, "A review on fault classification methodologies in power transmission systems: Part—I," *J. Electr. Syst. Inf. Technol.*, vol. 5, no. 1, pp. 48–60, May 2018, doi: 10.1016/j.jesit.2017.01.004.

[9] A. Dasgupta, S. Nath, and A. Das, "Transmission line fault classification and location using wavelet entropy and neural network," *Electr. Power Compon. Syst.*, vol. 40, no. 15, pp. 1676–1689, Oct. 2012.

[10] A. A. Majd, H. Samet, and T. Ghanbari, "k-NN based fault detection and classification methods for power transmission systems," *Protection Control Modern Power Syst.*, vol. 2, no. 1, pp. 359–369, 2017.

[11] S. Ekici, "Support vector machines for classification and locating faults on transmission lines," *Appl. Soft Comput.*, vol. 12, no. 6, pp. 1650–1658, Jun. 2012.

[12] S. A. Gafoor and P. V. R. Rao, "Wavelet based fault detection, classification and location in transmission lines," in *Proc. IEEE Int. Power Energy Conf.*, Nov. 2006, pp. 114–118.

[13] K. Jnaneswar, B. Mallikarjuna, S. Devaraj, D. S. Roy, M. J. B. Reddy, and D. K. Mohanta, "A real-time DWT and traveling waves-based multi-functional scheme for transmission line protection reinforcement," *Electr. Eng.*, vol. 103, no. 2, pp. 965–981, Apr. 2021, doi: 10.1007/s00202-020-01117-0.

[14] Y. S. Rao, G. R. Kumar, and G. K. Rao, "A new approach for classification of fault in transmission line with combination of wavelet multi resolution analysis and neural networks," *Int. J. Power Electron. Drive Syst.*, vol. 8, no. 1, p. 505, Mar. 2017.

[15] A. Harish and M. V. Jayan, *Inventive Communication and Computational Technologies* (Lecture Notes in Networks and Systems), vol. 145. Singapore: Springer, 2021, doi: 10.1007/978-981-15-7345-3.

[16] P. Gopakumar, M. J. B. Reddy, and D. K. Mohanta, "Transmission line fault detection and localisation methodology using PMU measurements," *IET Gener., Transmiss. Distrib.*, vol. 9, no. 11, pp. 1033–1042, Aug. 2015.

[17] M. Shafiullah and M. A. Abido, "S-transform based FFNN approach for distribution grids fault detection and classification," *IEEE Access*, vol. 6, pp. 8080–8088, 2018.

[18] A. Anand and S. Affijulla, "Hilbert–Huang transform based fault identification and classification technique for AC power transmission line protection," *Int. Trans. Electr. Energy Syst.*, vol. 30, no. 10, pp. 1–15, 2020.

[19] J. Upendar, C. P. Gupta, and G. K. Singh, "Statistical decision-tree based fault classification scheme for protection of power transmission lines," *Int. J. Electr. Power Energy Syst.*, vol. 36, no. 1, pp. 1–12, Mar. 2012, doi: 10.1016/j.ijepes.2011.08.005.

[20] A. G. Shaik and R. R. V. Pulipaka, "A new wavelet based fault detection, classification and location in transmission lines," *Int. J. Electr. Power Energy Syst.*, vol. 64, pp. 35–40, 2015, doi: 10.1016/j.ijepes.2014.06.065.

[21] A. Swetapadma and A. Yadav, "A novel decision tree regression-based fault distance estimation scheme for transmission lines," *IEEE Trans. Power Del.*, vol. 32, no. 1, pp. 234–245, Feb. 2017.

[22] P. K. Mishra, A. Yadav, and M. Pazoki, "A novel fault classification scheme for series capacitor compensated transmission line based on bagged tree ensemble classifier," *IEEE Access*, vol. 6, pp. 27373–27382, 2018.

[23] A. Jamehbozorg and S. M. Shahrtash, "A decision tree-based method for fault classification in double-circuit transmission lines," *IEEE Trans. Power Del.*, vol. 25, no. 4, pp. 2184–2189, Oct. 2010.

[24] J. C. A. Freire, A. R. G. Castro, M. S. Homci, B. S. Meiguins, and J. M. D. Morais, "Transmission line fault classification using hidden Markov models," *IEEE Access*, vol. 7, pp. 113499–113510, 2019.

[25] K. Chen, J. Hu, and J. He, "Detection and classification of transmission line faults based on unsupervised feature learning and convolutional sparse autoencoder," *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 1748–1758, May 2018.

[26] Y. Q. Chen, O. Fink, and G. Sansavini, "Combined fault location and classification for power transmission lines fault diagnosis with integrated feature extraction," *IEEE Trans. Ind. Electron.*, vol. 65, no. 1, pp. 561–569, Jan. 2018.

[27] S. Pandey, A. K. Srivastava, and B. G. Amidan, "A real time event detection, classification and localization using synchrophasor data," *IEEE Trans. Power Syst.*, vol. 35, no. 6, pp. 4421–4431, Nov. 2020.

[28] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189–215, 2020.

[29] J. Chorowski, J. Wang, and J. M. Zurada, "Review and performance comparison of SVM-and ELM-based classifiers," *Neurocomputing*, vol. 128, pp. 507–516, Mar. 2014.

[30] K. B. Swain, S. S. Mahato, and M. Cherukuri, "Expeditious situational awareness-based transmission line fault classification and prediction using synchronized phasor measurements," *IEEE Access*, vol. 7, pp. 168187–168200, 2019.

[31] W. Yuill, A. Edwards, S. Chowdhury, and S. P. Chowdhury, "Optimal PMU placement: A comprehensive literature review," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Jul. 2011, pp. 1–8.

[32] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.

[33] F. Thabtah, S. Hammoud, F. Kamalov, and A. Gonsalves, "Data imbalance in classification: Experimental evaluation," *Inf. Sci.*, vol. 513, pp. 429–441, Mar. 2020.

[34] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, 2006.

[35] G. Huang, G.-B. Huang, S. Song, and K. You, "Trends in extreme learning machines: A review," *Neural Netw.*, vol. 61, pp. 32–48, Jan. 2015, doi: 10.1016/j.neunet.2014.10.001.

[36] H. Zhang, "A new extreme learning machine optimized by particle swarm optimization," in *Proc. Int. Conf. Comput., Commun. Autom. (I3CA)*, no. 100, 2017, pp. 1–6.

[37] P. R. Bal and S. Kumar, "WR-ELM: Weighted regularization extreme learning machine for imbalance learning in software fault prediction," *IEEE Trans. Rel.*, vol. 69, no. 4, pp. 1355–1375, Dec. 2020.

[38] G. Huang, S. Song, J. N. D. Gupta, and C. Wu, "Semi-supervised and unsupervised extreme learning machines," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2405–2417, Dec. 2017.

[39] W. Zong, G.-B. Huang, and L. Chen, "Weighted extreme learning machine for imbalance learning," *Neurocomputing*, vol. 101, pp. 229–242, Feb. 2013, doi: 10.1016/j.neucom.2012.08.010.

[40] R. Eberhart and J. K. Sixth, "A new optimizer using particle swarm theory," *In: Proc. IEEE Symp. Micro Mach. Hum. Sci.* Nagoys, Japan, Oct. 1997, pp. 39–43.

[41] M. Clerc, "Particle swarm optimization," in *Proc. ICNN Int. Conf. Neural Netw.*, vol. 1, Nov. 2010, pp. 1942–1948.

[42] T. A. Universit. *Electric Grid Test Cases Library*. Accessed: Jul. 14, 2022. [Online]. Available: https://electricgrids.engr.tamu.edu/electric-grid-test-cases/ieee-39-bus-system/

[43] P. Demetriou, M. Asprou, J. Quiros-Tortos, and E. Kyriakides, "Dynamic IEEE test systems for transient analysis," *IEEE Syst. J.*, vol. 11, no. 4, pp. 2108–2117, Dec. 2015.

[44] F. Milano, "An open source power system analysis toolbox," *IEEE Trans. Power Syst.*, vol. 20, no. 3, pp. 1199–1206, Aug. 2005.

[45] M. Brown, M. Biswal, S. Brahma, S. J. Ranade, and H. Cao, "Characterizing and quantifying noise in PMU data," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, Jul. 2016, pp. 1–5.

[46] A. Bhandari, H. Yin, Y. Liu, W. Yao, and L. Zhan, "Real-time signal-to-noise ratio estimation by universal grid analyzer," in *Proc. Int. Conf. Smart Grid Synchronized Meas. Anal. (SGSMA)*, May 2019, pp. 1–6.

ANI HARISH (Senior Member, IEEE) received the B.Tech. degree in electrical engineering from Mahatma Gandhi University, Kerala, India, in 1997, and the M.Tech. degree in power systems from Kerala University, India, in 2000. She is currently pursuing the Ph.D. degree with APJ Abdul Kalam Technological University, Kerala.

Her research interests include wide area measurements, wide area monitoring and control of power systems, and artificial intelligence applications in power systems.

A. PRINCE (Senior Member, IEEE) received the B.Tech. degree in electrical engineering from Kerala University, India, in 1996, the M.Tech. degree in energy studies, in 2005, and the Ph.D. degree in power systems from the Indian Institute of Technology Delhi, Delhi, India, in 2012.

He has been in academics for 21 years. He is currently working as a Professor in electrical engineering with the Rajiv Gandhi Institute of Technology, Kerala, India. His research interests include power systems, renewable energy systems, wide area measurements, electric vehicles and signal processing applications in power systems, power electronics applications in power systems, and artificial intelligence applications in power systems.

Dr. Prince has contributed as a Reviewer of IEEE Transactions on Power Systems and IEEE Transactions on Industrial Electronics.

M. V. JAYAN received the B.Tech. degree in electrical engineering from Kerala University, India, in 1987, the M.E. degree in control systems from Bharathiar University, Coimbatore, India, in 2003, and the Ph.D. degree in high voltage engineering from the Indian Institute of Science, Bengaluru, India, in 2009.

He retired as a Professor in electrical engineering from the Government College of Engineering, Thrissur, Kerala, India, in 2021. His current research interests include high voltage engineering, solid pollution control, wide area monitoring and control of power systems, and power quality enhancement in grid integrated power systems.