A Synchro-phasor Assisted Optimal Features Based Scheme for Fault Detection and Classification

Homanga Bharadhwaj
Department of Computer Science
Indian Institute of Technology Kanpur
India
homangab@cse.iitk.ac.in

Avinash Kumar
Department of Electrical Engineering
Indian Institute of Technology Kanpur
India
kumarav@iitk.ac.in

Abheejeet Mohapatra
Department of Electrical Engineering
Indian Institute of Technology Kanpur
India
abheem@iitk.ac.in

***Abstract*—A novel and efficient methodology for comprehensive fault detection and classification by using synchrophasor measurement based variations of a power system is proposed. Presently, Artificial Intelligence (AI) techniques have been used in power system protection owing to the greater degree of automation and robustness offered by AI. Evolutionary techniques like Genetic Algorithm (GA) are efficient optimization procedures mimicking the processes of biological evolution that have been shown to perform better than their gradient based counterparts in many problems. We propose a combined GA and Particle Swarm Optimization (PSO) approach to find the optimal features relevant to our fault detection process. As is evidenced by recent advances in multi-modal learning, it has been shown that this combined approach yields a more accurate feature optimization than that obtained by a single meta-heuristic. A systematic comparison of Artificial Neural Network (ANN) and Support Vector Machine (SVM) based methods for fault classification using the identified optimal features is presented. The proposed algorithm can be effectively used for real time fault detection and also for performing postmortem analysis on signals. We demonstrate its effectiveness by simulation results on real world data from the North American SynchroPhasor Initiative (NASPI) and signal variations from a test distribution system.**

***Index Terms*- Genetic algorithm (GA), particle swarm optimization (PSO), phasor measurement unit (PMU), optimal features, fault detection and classification, support vector machine (SVM), artificial neural network (ANN)**

I. Introduction

A fault is defined as an abnormal condition or defect in one or many power system components at a time which may lead to partial or complete failure of the power system, if not detected and cleared quickly and judiciously. Protective relays are used to trip circuit breakers in a power system in the event of occurrence of a fault. Specifying an absolute threshold for the values of voltage and current signals to detect a fault is not faithful as this does not take into account the various statistical measures of the signals that change in the event of a fault [1]. Analyzing signals from a Phasor Measurement Unit (PMU) has its challenges due to the non-stationary nature of the signals obtained, which lack the fundamental frequency [2]. Identification of optimal features prior to classification is important so that the Machine Learning (ML) pipeline does not produce errors due to outliers in irrelevant features and also to reduce the computational complexity (of having to compute a large number of irrelevant features) during testing. In the event of occurrence of a fault, one or many signal(s) in the power system such as bus voltages, line currents, power, and frequency have abrupt variations. Hence, the optimal or most suitable feature selection needs to be known from these variations. Also, the feature selection process should take into account all fluctuations and differences in the above phasors so that the identified features appropriately represent the signal under all conditions. [3].

Fault detection approaches [4] are mainly model based, the signal based, knowledge-based and hybrid methods which combine the previous three. With the massive explosion of big data analysis and data mining techniques [5], [6], [7] in recent times, the knowledge-based approaches have been highly popular even in the field of fault diagnosis. The signal or data required for the knowledge-based approaches are readily available from the Remote Terminal Units (RTUs) placed in Supervisory Control And Data Acquisition (SCADA) systems and also from PMUs [8], [9], [10]. This also makes the collection of bulk historical data a convenient endeavor. ML techniques [11], [12] that classify the regions of an electrical signal as being faulty or otherwise have been considerably researched upon. Selection of optimal features [13]–[15] to remove redundant and irrelevant features before classification in the fault detection process have mainly employed standard techniques like Principal Components Analysis (PCA), Independent Components Analysis (ICA) or standalone evolutionary techniques like GA, PSO, Genetic Programming (GP) [16], [17]. Hence, they are susceptible to typical pitfalls like requiring a lot of training data (PCA, ICA), not converging for a large number of iterations (GA, GP) and settling in a local optima (PSO) leading to the identification of sub-optimal features [16]. The method in [11] applies PCA to the signals obtained via Wavelet Transform to reduce the data dimensionality. This only achieves removal of highly correlated features and does not guarantee the selection of optimal features which is necessary for successful detection and classification of faults. The method in [3] directly uses the signals from Discrete Wavelet Transform (DWT) without reducing the dimensionality or optimizing for features before fault classification. Another drawback is that the method attempts to directly classify faults without first detecting if there is a fault.

To mitigate these issues, a novel methodology that explicitly optimizes for features and then employs fault detection fol-

lowed by fault classification has been proposed in this paper. Based on past studies [18]–[20] DWT is applied to extract the approximation and detail coefficients from the signals as it is more convenient to work with them rather than the bulky original phasors. The proposed methodology has two key phases.

- In the first phase, the optimal feature selection from the variations of the various metered power system signals such as bus voltage magnitude, phase angle, power, and frequency is done through the use of a combined GA-PSO approach. The objective during the feature selection process is the minimization of the Mean Square Error (MSE) on the validation set of the data set. In this phase, GA is used on a certain fraction of the total population while PSO is used to the other fraction in an iteration of this approach. This approach in the first phase is expected to be highly robust and reliable because evolutionary techniques are known to converge to the global optima (and not get trapped in local optima unlike gradient-based techniques) [21], [22]. Also, the proposed methodology can be used for real-time fault monitoring as the first phase (which is expected to be slow due to combined GA-PSO approach) plays a pivotal role only during the training phase.
- SVM and ANN are used in the second phase of the methodology for fault classification and detection, once the optimal features have been identified. The novelty in this phase is the adoption of two separate pipelines for fault classification and fault detection. It is important to note that it is not necessary to use supervised ML algorithms like SVMs and ANNs in this phase. After identification of the optimal features by the combined GA-PSO approach, an unsupervised or semi-supervised method as in [23] for fault detection can also be used. So, the methodology here is of broader scope as the authors in [23] do not employ feature selection or optimization. Also, since feature optimization occurs offline during the training phase, the proposed methodology is expected to be extremely efficient for use in real time fault detection and classification in power systems.

The rest of the paper is organized as follows. Section II describes the basics of the various techniques used in our model. Section III presents a detailed description of the proposed optimal feature based classification methodology for fault detection and classification in power systems. Section IV discusses the simulation results obtained by testing the proposed methodology on multiple datasets. Finally, section V concludes the paper with a summary of the key contributions.

II. Background of the Proposed Approach

A power system fault is characterized by abnormal variations in current, amplitude, phase angle, frequency, real power, and apparent power. Fault detection and isolation is an essential part of keeping the power system safe and usable. Doing this manually is not feasible in the current scenario of various intricate smart grids and distribution networks. Here, we develop a strategy for fault detection and classification using various AI techniques that are guaranteed to be more ubiquitously usable, robust and requires minimal human intervention. We first describe in brief the procedures of DWT, GA, PSO, SVM and ANN which are used in our method and then illustrate our proposed approach in the next section.

A. Discrete Wavelet Transform (DWT)

Hence, the higher level coefficients are obtained. In this work, we employ only the d3 coefficients of the signals [24] obtained from PMUs. The use of d3 coefficients is motivated by the paper [23] which argues for the efficacy of d3 coefficients being empirically the most optimal for optimizing features for fault detection.

B. Feature Engineering and Selection

The process of feature engineering followed by feature selection is referred to as feature extraction. The architecture of features requires domain knowledge of experts and is essential for making any learning algorithm work. In our case, it is reasonable to consider significant statistical variables like 'rate of change', standard deviation and energy as the possible relevant features [24]. Having engineered a lot of features, a selection mechanism needs to be employed for selecting the features optimal to the given learning process [3].

In this work, we employ two meta heuristic algorithms, namely Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) for feature selection. In GA, a population of satisfactory solutions (called individuals or chromosomes) to an optimization problem is evolved to a better solution. Typically, a fitness function is used for evaluating the solution space in each generation [25]. In PSO, a population or swarm of particles is initialized randomly and then moved around in the search space by following certain pre-defined formulae [16]. The movements of each particle are typically motivated both by its own best-known position and by the swarm's best-known position.

III. The Proposed Approach

We employ our proposed approach as a scheme of feature optimization on a validation set via a combined GA-PSO approach followed by the fault detection and classification via ANN and SVM. The benefit of using evolutionary heuristic approaches for feature optimization on the validation set is that only the features most suitable for our particular task will be identified, which are not necessarily the entire set of least-correlated features (as identified by dimensionality reduction approaches like PCA and ICA). Algorithm 1 summarizes the basic workflow of our methodology. Since in the occurrence of a fault, the dynamics of every variable namely voltage, angle, power, frequency, etc. behave abruptly, so we use features extracted from all these variations. Hence our method is 'comprehensive' in the sense that it considers variations in multiple signals for fault detection.

The first step is the extraction of features from our data. The data we have is from a PMU, and there are variations of

Data: Given signals for variation in power, angle, voltage, and frequency
for *every signal (V,P,f,A)* **do**
  Extract the d3 coefficients by applying DWT
end
Set the value of base sampling rate l, **for** *r=l;r* $\leq$ *6l;r =r+l* **do**
  for *every signal G =[V,P,f,A]* **do**
    Sample at the rate R
    Keep 10% of the total samples, say n_1 for validation of optimal features
    The remaining n_2 samples will be used for training of the classifiers
    Extract the features ΔG, $\sigma(G)$ and $E(G)$
  end
  Set the objective function Q_1 for GA/PSO to be the mean square error in validation set, $Q_1 = \frac{\sum_{i=1}^{n_1}(y(i)-\hat{y}(i))^2}{n_1}$ where y is the actual output and $\hat{y}$ is the output of the ANN classifier **while** *Stopping criteria of GA/PSO not reached* **do**
    Using the current optimal features, train the ANN/SVM on the training samples (n_2 samples)
    for *i=1; i* $\leq n_1$*; i++* **do**
      Determine the output $\hat{y}(i)$
    end
    Evaluate $Q_1 = \frac{\sum_{i=1}^{n_1}(y(i)-\hat{y}(i))^2}{n_1}$
  end
  Set the objective function Q_2 for GA/PSO to be the mean square error in validation set, $Q_2 = \frac{\sum_{i=1}^{n_1}(y(i)-\hat{y}(i))^2}{n_1}$ where y is the actual output and $\hat{y}$ is the output of the SVM classifier
  while *Stopping criteria of GA/PSO not reached* **do**
    Using the current optimal features, train the SVM/ANN on the training samples (n_2 samples)
    for *i=1; i* $\leq n_1$*; i++* **do**
      Determine the output $\hat{y}(i)$
    end
    Evaluate $Q_2 = \frac{\sum_{i=1}^{n_1}(y(i)-\hat{y}(i))^2}{n_1}$
  end
end

Algorithm 1: Stepwise algorithm for the training methodology of fault detection and classification

frequency (f), voltage (V), angle (θ) and power (P) in the data. The Daubechies DWT of each signal is performed, and the d3 coefficients are recorded. Since it is a very standard technique, we do not elaborate its details here, but refer the reader to [26] for a comprehensive tutorial. Henceforth, the signals referred to are the d3 coefficients of the respective analog signals. The simplistic approach of defining a fault as a region where individual variables exceed a certain threshold can give erroneous results in many situations. So, hard encoding is not an option, and one must devise an efficient ML work flow. The knowledge of what should be the relevant features that define a fault can come only from experience, and hence to eliminate human intervention, we train our model to 'learn' the best subset of features. Before feature extraction, the data is sampled at a suitable rate. The base rate is said l. We run experiments for different values of sampling rate, namely l, $2l$, $3l$, $4l$, $5l$ and $6l$. We henceforth denote the sampling rate by R. The optimal sampling rate cannot be known a priori, and so it is a hyper-parameter of our technique that will require tuning.

For feature extraction, the rate of change of a variable (voltage, angle, etc.), energy and variance of each sample are calculated. From previous research [Cite - some papers of Seethalakshmi / SN Singh], these features are crucial for deploying systems to detect and classify faults. However, not all of them may be relevant for a particular use-case, and hence we select optimal features from that set for particular tasks. These are the initial features employed in our technique which for a given sampled signal G are given as

$$\Delta G = G(n) - G(n-R) \tag{1}$$

$$\sigma(G) = \sqrt{Variance[G(n-R:n)]} \tag{2}$$

$$E(G) = \sum_{i=n-R}^{n} |G(i)|^2 \tag{3}$$

Here, $G = [\theta, P, f, V]$, n is the sample count and R, as mentioned above is the sampling rate in samples/cycle. As the system learns, it will identify the features which are optimal to the process (the features employing which give the least error on the validation dataset). During the test, only the optimal features are used in the model. The selection of optimal features [3] has been implemented by a combination of GA and PSO. Let the population size initialized be $4N$. Now, in each iteration (or generation), $2N$ individuals are randomly chosen to be evolved through GA, and the other $2N$ particles are evolved through PSO. For the final $4N$ particles, the above procedure is repeated for the next overall iteration. Algorithm 2 summarizes the combined GA-PSO approach succinctly.

As we demonstrate in the simulation results, this method is indeed better than just using GA or PSO because a form of multi-modal learning is in effect here [20]–[24], [26]. Learning two alternate representations of the data distribution through two techniques - GA and PSO [21], allows us to capture the peculiarities of the data and its modalities well. Hence, multi-modal learning is a very strong modeling technique for learning, and we use it in our algorithm to identify the optimal features. Also, the model is less susceptible to get trapped in local optima because both these evolutionary algorithms are not gradient based. These meta-heuristic algorithms offer effective convergence to the global optima as opposed to gradient-based methods which are highly susceptible to settle for local optima. The usual technique of feature selection is to optimize an objective function such as mutual information, mutual correlation or cross entropy. What these methods in effect achieve is to yield a set of features which are least correlated, but not necessarily best for the given classification or regression process. The present approach considers the error

over a validation set (taken as 10% of the training dataset) as the fitness function and optimizes it.

Data: Randomly initialize $4N$ chromosomes each of length n
Initialize the fitness of all chromosomes to 0
while *Max number of iterations not reached* **do**
  I. Randomly select $2N$ individuals from the population (One generation for GA)
  for $i = 1; i \leq N; i++$ **do**
    1) Selection - Select two of the N chromosomes with the best fitness
    2) Crossover - Perform uniform crossover to generate a offspring
    3) Mutation - With a probability of 30% select two random bits of the offspring and flip their values
  end
  Evaluate the fitness value of the N parents and N offsprings on $Q = \frac{\sum_{i=1}^{n_1}(y(i)-\hat{y}(i))^2}{n_1}$
  II. Select the remaining $2N$ particles in the population (One iteration for PSO)
  for $i = 1; i \leq 2N; i++$ **do**
    Update the particle's velocity as $\mathbf{v_i} = \eta\mathbf{v_i} + \zeta_p k_p(\mathbf{p_i} - \mathbf{x_i}) + \zeta_g k_g(\mathbf{g} - \mathbf{x_i})$ and position as $\mathbf{x_i} = \mathbf{x_i} + \mathbf{v_i}$
    if $h(\mathbf{x_i}) < h(\mathbf{p_i})$ **then**
      $\mathbf{p_i} = \mathbf{x_i}$
      if $h(p_i) < h(x_i)$ **then**
        $\mathbf{g} = \mathbf{p_i}$
      end
    end
    Evaluate the fitness function Q for the current particle
  end
  Now, combine all the above outputs to obtain $4N$ individuals in the total population
end

Algorithm 2: The combined GA-PSO based feature selection

For the GA phase of the combined GA-PSO approach, the chromosomes consist of n binary digits which denotes the total number of features extracted from data. A random population of chromosomes is initially generated, the fitness of which is measured by checking how the overall classification process performs on it. Each chromosome contains an indication of which features are to be used in the current computation. The chromosomes in each generation which yield the least classification error on the validation dataset are suitably passed on to the next generation. Offsprings are produced by crossover and mutation of the previously chosen best-fit parents. The technique of uniform crossover is employed, where bits are randomly copied from either the first or second parent. For mutation, two bits are selected at random, and their values are flipped. For the PSO phase, a random population of particles is initialized. There are as many dimensions in the search space as the total number of features in consideration. Each particle in the swarm represents a subset of the overall features. The swarm is evolved to obtain a global best position and also individual best positions of the particles, as described in section II. The classification performance of each particle (feature subset) is evaluated on the validation set, as an inner loop in the training process [17]. In this way, after multiple iterations, the GA-PSO algorithm finally converges to an optimal individual, which is a chromosome of ones and zeros. The ones indicate the desired optimal features.

The objective function for feature selection (combined GA-PSO approach as in algorithm 2) is $Q = \frac{1}{n}\sum_{i=1}^{n_1}(y(i)-\hat{y}(i))^2$, where $y(i)$ is the correct output of the i^{th} sample in the validation set and $\hat{y}(i)$ is the model output with the current subset of features. This is evaluated to identify the best fit individuals in every iteration. After the selection of optimal features, we employ two different models for fault classification and fault location. The first model, which we call SVM-ANN, use a SVM for classification of faults (the different types of fault are shown in Table I) and an ANN for the location of faults, which is essentially a regression task. Both the SVM and ANN use the same set of optimal features previously identified by the combined GA-PSO approach. The second model, which we call ANN uses two ANNs, one for classification of faults and the other for regression of fault location. Both the ANNs use the same set of optimal features generated previously by the combined GA-PSO approach. 70% of the samples are selected for training our models and a comparative study of the two models is presented. The ANN used is a double hidden layer architecture with ten nodes in each hidden layer. Linear SVM, Quadratic SVM, Cubic SVM and RBF (Radial Basis Function) kernel SVM were trained for the SVM, but we found RBF kernel to work the best on both the datasets in Section IV. Hence the results reported are by using the RBF kernel in SVM.

TABLE I
CORRESPONDENCE OF FAULT TYPES TO CLASSIFIER OUTPUT. THE FAULT TYPE SYMBOLS HAVE THEIR STANDARD MEANING AS DESCRIBED IN PREVIOUS RESEARCH [CITE –COMBINED FAULT LOCATION AND CLASSIFICATION FOR POWER TRANSMISSION LINES FAULT DIAGNOSIS WITH INTEGRATED FEATURE EXTRACTION]

Index	0	1	2	3	4	5	6
Fault Type	No Fault	A-G	B-G	C-G	A-B	B-C	A-C
Index	7		8		9		10
Fault Type	A-B-G		B-C-G		A-C-G		A-B-C

After training the models, a comparative analysis of the SVM-ANN and ANN techniques on the 30% test data is performed and many interesting observations are reported and analyzed. Now, the rate of sampling which is a hyper parameter for our process is tuned. The rate is varied, and the resulting classification accuracy is observed. We expect the accuracy to decrease when the rate is very high because then sufficient time would not have elapsed for detection of an event. In this case, even faulty samples would be classified as non faulty. On the other hand, the accuracy is also expected to

decrease when the rate is meager because then the classifier would not have sufficient data samples to be trained efficiently and hence falter. The sampling rate which gives the best result in training is chosen and applied to the test signals. The optimal features and the trained models can be used for predicting faults in real time or on a test dataset for postmortem analysis.

The most commonly cited argument against evolutionary meta heuristics algorithms like GA, PSO, etc. is that they are slow to learn. In our case, we use them only during the training phase, for selecting an optimal subset of features. While testing the model, the optimal features identified during training are used for fault location and classification (using SVM-ANN or ANN models). Hence, our model is suitable for real-time fault monitoring.

IV. Simulation Results

A. Case I (NASPI dataset)

This section presents the results of implementing our algorithm on real world PMU data from NASPI (North American SynchroPhasor Initiative) [27]. The variations in voltage, angle, power and frequency obtained from a PMU for a duration of 10 minutes (as shown in Fig. 1) are used for testing our algorithm. DWT is applied to all signals using Daubechies wavelets and $d3$ coefficients are recorded, as shown in Fig. 2.

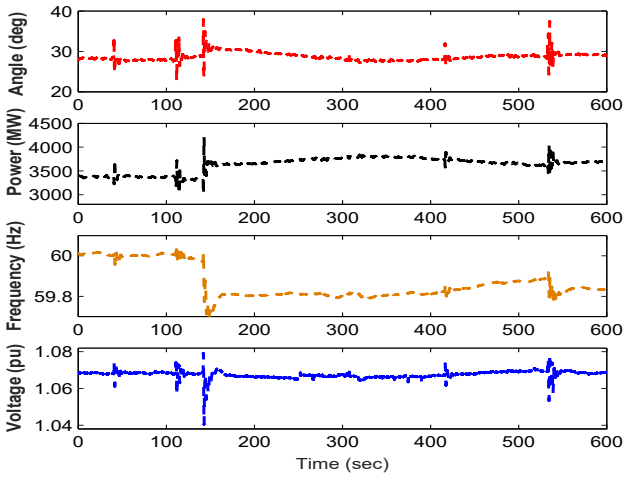

Fig. 1. Analog signals (From NASPI dataset)

1) Feature Selection: For feature selection, the features are arranged as shown on the right. The numbering corresponds to the gene-location of each feature in the chromosome. For example, ΔP is the third gene in the chromosome.

Index	1	2	3	4	5	6
Feature	$\Delta\theta$	Δf	ΔP	ΔV	$\sigma(\theta)$	$\sigma(V)$
Index	7	8	9	10	11	12
Feature	$\sigma(P)$	$\sigma(P)$	$E(\theta)$	$E(f)$	$E(P)$	$E(V)$

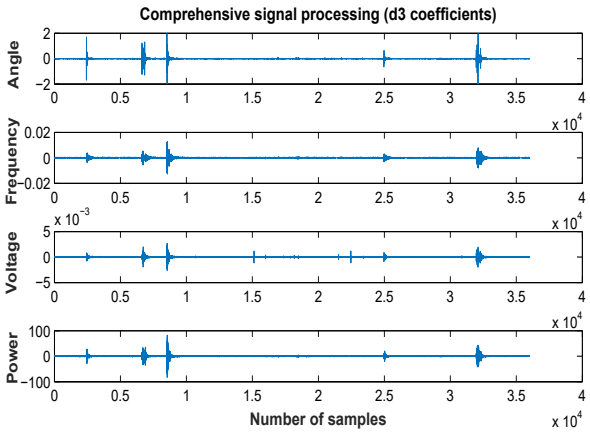

Fig. 2. Daubechies DWT d3 coefficients for signals in Fig. 1

So, the initial population contains chromosomes (particles in case of PSO) which are strings of randomly assigned 0 or 1 for each of the twelve genes. In every iteration of the combined GA-PSO algorithm, the chromosomes are changed. Every chromosome is evaluated on the fitness function mentioned in section III. This means that for each in every iteration, classification using the current features (indicated by 1's in gene locations) must be performed (i.e., the model is trained) and the error on the validation set is evaluated. The best-fit individual we obtain for the present NASPI dataset, wherein each sample has 20 data points is as shown.

0	0	1	0	0	1	0	1	0	1	1	0

It means that of the twelve features, only five are optimal for the present problem at the given sample size. Now, we present a detailed tabulation of variation of optimal features with sample size in Table II (note that the base rate l corresponds to 10 data points per sample). It can be seen that the above best-fit individual is for a sample size of $2l$ in Table II.

TABLE II
Optimal Features with the combined GA-PSO technique

Sample size	Best Fit Individual											
l	1	0	1	0	1	0	1	1	0	1	1	1
$2l$	0	0	1	0	0	1	0	1	0	1	1	0
$3l$	0	1	1	0	0	1	0	1	0	1	0	1
$4l$	0	0	1	0	1	1	0	1	0	0	0	1
$5l$	1	1	1	0	0	1	0	0	1	1	0	0
$6l$	0	1	1	0	0	1	0	1	0	1	0	1

As shown in Table III, neither GA nor PSO uniformly gives better test classification accuracy for all sample sizes. GA usually takes larger iterations to converge to an optimal solution than PSO. When the stopping criteria are the number of iterations, then GA may perform worse typically when the sample size is small (number of samples are high) [11]. The proposed GA-PSO method in a sense combines the best of

TABLE III

% CLASSIFICATION ERROR IN FAULT CLASSIFICATION FOR CASES OF NO FEATURE OPTIMIZATION (ALL FEATURES), GA-BASED FEATURE OPTIMIZATION (GA), PSO BASED FEATURE OPTIMIZATION (PSO) AND THE PROPOSED COMBINED GA-PSO BASED FEATURE OPTIMIZATION FOR CASE I

Sample Size	All Features				GA				PSO				Combined GA-PSO			
	ANN		SVM-ANN		ANN		SVM-ANN		ANN		SVM-ANN		ANN		SVM-ANN	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
l	0.10	0.24	0.04	0.17	0.09	0.25	0.08	0.17	0.09	0.22	0.08	0.15	0.08	0.20	0.02	0.13
2l	0.08	0.21	0.02	0.15	0.08	0.22	0.05	0.15	0.07	0.19	0.08	0.13	0.06	0.18	0.01	0.11
3l	0.29	0.43	0.04	0.21	0.24	0.39	0.15	0.18	0.22	0.32	0.20	0.18	0.20	0.27	0.02	0.16
4l	0.63	0.95	0.03	0.23	0.57	0.84	0.22	0.19	0.55	0.84	0.51	0.18	0.51	0.81	0.02	0.16
5l	0.72	1.3	0.05	0.25	0.69	1.1	0.21	0.19	0.70	1.25	0.64	0.22	0.62	1.1	0.03	0.18
6l	0.92	1.5	0.05	0.25	0.85	1.3	0.23	0.20	0.87	1.48	0.85	0.19	0.80	1.1	0.03	0.18

TABLE IV

% CLASSIFICATION ERROR IN FAULT CLASSIFICATION FOR CASES OF NO FEATURE OPTIMIZATION (ALL FEATURES), GA-BASED FEATURE OPTIMIZATION (GA), PSO BASED FEATURE OPTIMIZATION (PSO) AND THE PROPOSED COMBINED GA-PSO BASED FEATURE OPTIMIZATION FOR CASE II

Sample Size	All Features				GA				PSO				Combined GA-PSO			
	ANN		SVM-ANN		ANN		SVM-ANN		ANN		SVM-ANN		ANN		SVM-ANN	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
l	0.03	0.09	0.03	0.11	0.09	0.25	0.08	0.17	0.09	0.21	0.08	0.13	0.02	0.06	0.02	0.07
2l	0.03	0.07	0.02	0.09	0.08	0.22	0.05	0.15	0.07	0.17	0.08	0.11	0.01	0.04	0.01	0.06
3l	0.09	0.08	0.03	0.10	0.24	0.39	0.15	0.18	0.22	0.32	0.20	0.15	0.01	0.05	0.02	0.06
4l	0.12	0.19	0.03	0.18	0.57	0.84	0.22	0.19	0.55	0.84	0.51	0.17	0.09	0.15	0.03	0.11
5l	0.21	0.29	0.04	0.25	0.69	1.1	0.21	0.19	0.70	1.25	0.64	0.21	0.14	0.24	0.03	0.23
6l	0.25	0.65	0.05	0.41	0.85	1.3	0.23	0.20	0.87	1.48	0.85	0.19	0.32	0.61	0.04	0.34

both worlds and yields lesser test error that both GA and PSO alone. Multi-modal learning, which underlies the combined GA-PSO method enables it to be more robust by modeling greater stochasticity in the underlying distribution of features.

2) Training of the SVM: This section describes the training of SVM for the sample size of l. As mentioned in section III, we tested our model using polynomial kernels and RBF kernel in the SVM. The regularization parameter C of SVM requires tuning. Also, the parameter a of a polynomial kernel and parameter σ of the RBF kernel needs to be appropriately set. A five fold cross validation scheme is used to perform grid search for the hyper parameters, C, a and σ, in the range $[0.001, 5000]$, $[0.0002, 40000]$ and $[0.0001, 30000]$, respectively. Although in the present case, the results obtained by using all the four kernels are comparable, but in the case of RBF kernel, the test accuracy is slightly higher.

3) Training of the ANN: For the ANN, we use 10 nodes each in the two hidden layers and holdout cross validation. The weights are randomly initialized. The only hyper parameter for an ANN is the number of nodes in the hidden layers, and hence training it is comparatively more convenient.

4) Results: For every sample size, our model is tested with ANN and SVM classifiers with the methodology described in Section III. The optimal features for each sample size are used in the respective classification process. Table III shows the results of fault classification using various schemes for feature optimization and also for the scheme of no feature optimization (i.e., by using all the features).

This section describes the training of SVM for the sample size of l. As mentioned in section III, we tested our model using polynomial kernels and RBF kernel in the SVM. The regularization parameter C of SVM requires tuning. Also, the parameter a of a polynomial kernel and parameter σ of the RBF kernel needs to be appropriately set. A five fold cross validation scheme is used to perform grid search for the hyper parameters, C, a and σ, in the range $[0.001, 5000]$, $[0.0002, 40000]$ and $[0.0001, 30000]$, respectively. Although in the present case, the results obtained by using all the four kernels are comparable, but in the case of RBF kernel, the test accuracy is slightly higher.

5) Training of the ANN: For the ANN, we use 10 nodes each in the two hidden layers and holdout cross validation. The weights are randomly initialized. The only hyper parameter for an ANN is the number of nodes in the hidden layers, and hence training it is comparatively more convenient.

6) Results: For every sample size, our model is tested with ANN and SVM classifiers with the methodology described in Section III. The optimal features for each sample size are used in the respective classification process. Table III shows the results of fault classification using various schemes for feature optimization and also for the scheme of no feature optimization (i.e., by using all the features).

B. Case II (Data from a Distribution System)

1) Generation of the training data: A Real Time Digital Simulator (RTDS) platform installed in IIT Kanpur was used to obtain the dataset for the system shown in Fig. 5 whose detailed data specifications can be found in the supplementary document [28]. The dataset consists of 36000 data points spread evenly over a duration of $10.8s$. Faults are generated at multiple locations, and the duration of successive faults is gradually increased to capture a diverse training scenario and to demonstrate the robustness of our proposed algorithm in detecting the faults. As in Case I, the base sample size is taken to be $l = 10$, and results are generated for six different

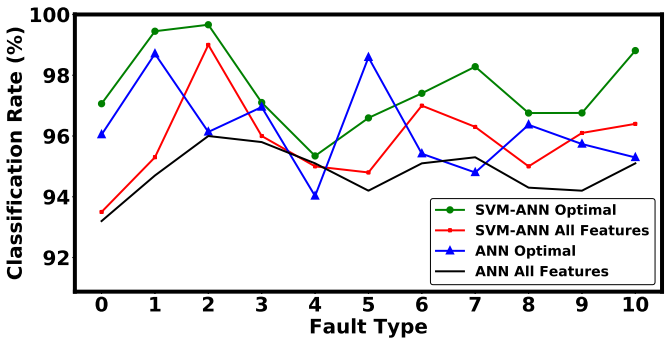

Fig. 3. Variation of classification accuracy for the four variants of the architecture (Case I). The reported values have been averaged over all the six cases of the sample sizes.

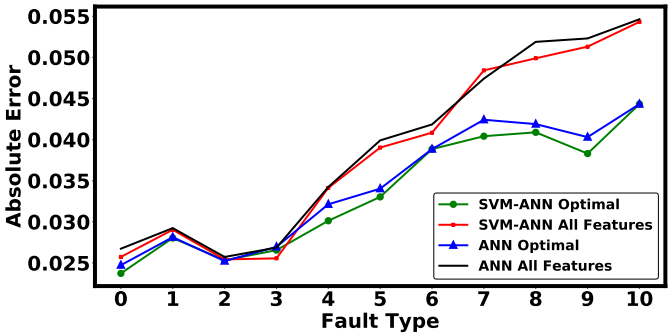

Fig. 4. Variation of absolute error in fault location for the four variants of the architecture (Case I). The reported values have been averaged over all the six cases of the sample sizes.

sample sizes. Daubechies DWT is applied before sampling as in Case I.

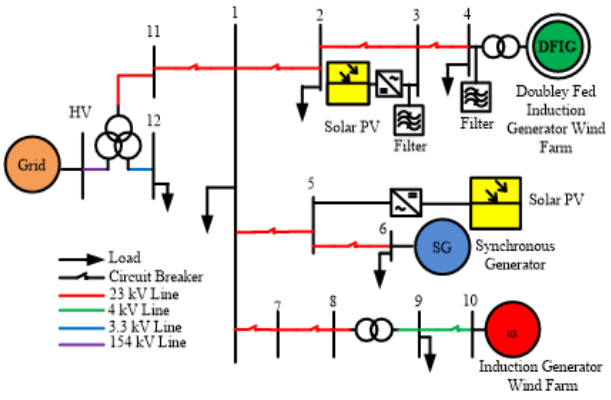

Fig. 5. Single line diagram of the distribution system

2) Test results: Table IV Justifies our claim that the combined GA-PSO approach yields better features which produce higher test accuracy. A similar trend of test accuracies as in Case I is observed here, making it concrete that improved test accuracy is not specific to a particular dataset. As Table IV shows, the least test errors are obtained for sample sizes $2l$ and $3l$ while the test errors for all sample sizes reduce after feature optimization as proposed in our proposed algorithm. Fig. 6 and Fig. 7 show the comparison of the two models using the combined GA-PSO optimized features and no optimization for fault classification and fault location, respectively. We observe again that feature optimization using the GA-PSO approach leads to better performance across fault types. Since, we observe this performance improvement in both Case I and Case II, i.e., across two different datasets, we can conclude that the employed method is indeed robust and not biased towards a particular dataset or a particular type of fault. Owing to space constraint, we omit the detailed discussions as in Case I.

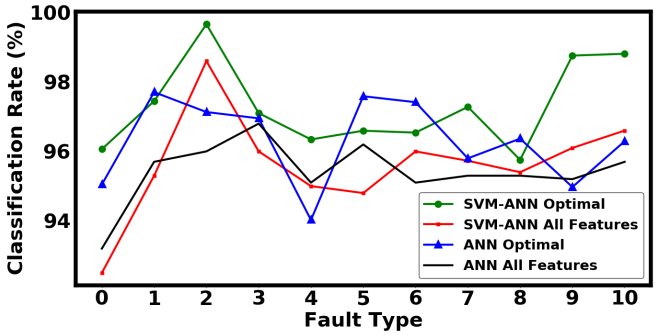

Fig. 6. Variation of classification accuracy for the four variants of the architecture (Case II). The reported values have been averaged over all the six cases of the sample sizes.

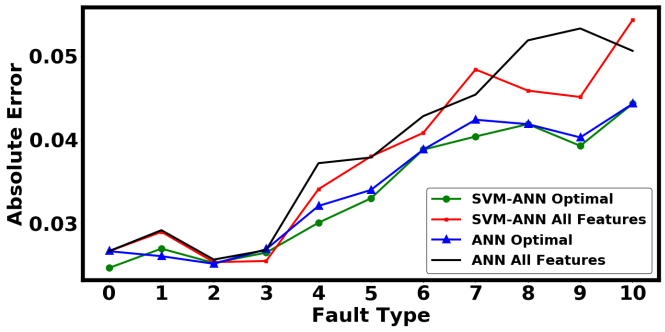

Fig. 7. Variation of absolute error in fault location for the four variants of the architecture (Case II). The reported values have been averaged over all the six cases of the sample sizes.

C. Analysis of Training and Prediction time

TABLE V
TRAINING AND PREDICTION TIME OF THE PROPOSED APPROACH FOR THE TWO SIMULATION EXPERIMENTS

	Time for Training (in hours)			Time for Prediction (in milliseconds)
	Feature Optimizer	SVM	ANN	
NASPI Dataset (A)	2.3	0.7	1.8	0.03
Data from PMU (B)	2.1	0.5	1.5	0.02

Table V illustrates the time required for offline training and real-time predictions. The training time although is high, this

is not a concern for deploying the model because testing can be done in real-time. Since the predictions times are very low, the real-time prediction can be done efficiently.

V. CONCLUSION

In this paper, an innovative strategy for fault detection and classification using optimal features based classifier is proposed. An approach combining GA and PSO is used for selecting the optimal features for the classification problem. The method involves applying GA to a fraction of the total population and applying PSO to the other fraction, in every iteration. As shown in the simulation results, the performance after feature optimization is much better than that obtained by using all the features for the SVM-ANN or ANN based classifiers. Also, the test error is lesser for our combined GA-PSO approach than just GA or PSO, indicating that the model has the lesser susceptibility to getting trapped in local optima. The method proposed is comprehensive in the sense that it considers variations in a number of phasors (voltage, power, frequency, and angle) for feature engineering. Finally, as we argued in the paper, this method is perfectly suitable both for real time fault detection and also post-mortem analysis of signals.

ACKNOWLEDGMENT

The authors would like to thank the DST/Indo-US Science and Technology Forum (IUSSTF), New Delhi, India and NTPC NETRA, Noida, India for providing financial support to carry out this research work under projects IUSSTF/EE/2017282, DST/EE/2018174 and NTPC/EE/2016288.

REFERENCES

[1] S. S. Negi, N. Kishor, K. Uhlen, and R. Negi, "Event detection and its signal characterization in pmu data stream," *IEEE Trans. Ind. Inf.*, vol. PP, no. 99, pp. 1–1, 2017.

[2] M. K. Jena, B. K. Panigrahi, and S. R. Samantaray, "A new approach to power system disturbance assessment using wide-area postdisturbance records," *IEEE Trans. Ind. Inf.*, vol. 14, no. 3, pp. 1253–1261, 2018.

[3] B. K. Panigrahi and V. R. Pandi, "Optimal feature selection for classification of power quality disturbances using wavelet packet-based fuzzy k-nearest neighbour algorithm," *IET Gener. Transm. Distrib.*, vol. 3, no. 3, pp. 296–306, March 2009.

[4] X. Dai and Z. Gao, "From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis," *IEEE Trans.Ind. Inf.*, vol. 9, no. 4, pp. 2226–2238, Nov. 2013.

[5] Morsi Ibrahim Walid G., Sami Alshareef and Saurabh Talwar , "Smart multi-purpose monitoring system using wavelet design and machine learning for smart grid applications," Patent US 20 160 124 031A1, May 05, 2016.

[6] Z. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques Part I: Fault diagnosis with model-based and signal-based approaches," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3757–3767, June 2015.

[7] E. Casagrande, W. L. Woon, H. H. Zeineldin, and N. H. Kan'an, "Data mining approach to fault detection for isolated inverter-based microgrids," *IET Gener. Transm. Distrib.*, vol. 7, no. 7, pp. 745–754, 2013.

[8] J. D. L. Ree, V. Centeno, J. S. Thorp, and A. G. Phadke, "Synchronized phasor measurement applications in power systems," *IEEE Trans. Smart Grid*, vol. 1, no. 1, pp. 20–27, June 2010.

[9] J. Kim, B. Lee, S. Han, J. H. Shin, T. Kim, S. Kim, and Y. Moon, "Study of the effectiveness of a korean smart transmission grid based on synchro-phasor data of k-wams," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 411–418, March 2013.

[10] M. Khan, P. M. Ashton, M. Li, G. A. Taylor, I. Pisica, and J. Liu, "Parallel detrended fluctuation analysis for fast event detection on massive pmu data," *IEEE Trans. Smart Grid*, vol. 6, no. 1, pp. 360–368, Jan 2015.

[11] J.-A. Jiang, C.-L. Chuang, Y.-C. Wang, C.-H. Hung, J.-Y. Wang, C.-H. Lee, and Y.-T. Hsiao, "A hybrid framework for fault detection, classification, and location Part II: Implementation and test results," *IEEE Trans. Pow. Del.*, vol. 26, no. 3, pp. 1999–2008, July 2011.

[12] M. Marseguerra and E. Zio, "Monte carlo simulation for model-based fault diagnosis in dynamic systems," *Reliability Engineering & System Safety*, vol. 94, no. 2, pp. 180–186, 2009.

[13] B. Xue, M. Zhang, and W. N. Browne, "Particle swarm optimization for feature selection in classification: A multi-objective approach," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1656–1671, Dec. 2013.

[14] I.-S. Oh, J.-S. Lee, and B.-R. Moon, "Hybrid genetic algorithms for feature selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 11, pp. 1424–1437, 2004.

[15] S. Raza, H. Mokhlis, H. Arof, K. Naidu, J. A. Laghari, and A. S. M. Khairuddin, "Minimum-features-based ANN-PSO approach for islanding detection in distribution system," *IET Renewable Power Gener.*, vol. 10, no. 9, pp. 1255–1263, Oct. 2016.

[16] K. O. Jones, "Comparison of genetic algorithm and particle swarm optimization," in *Proceedings of the International Conference on Computer Systems and Technologies*, 2005, pp. 1–6.

[17] M. A. Rodriguez-Guerrero, A. Y. Jaen-Cuellar, R. D. Carranza-Lopez-Padilla, R. A. Osornio-Rios, G. Herrera-Ruiz, and R. d. J. Romero-Troncoso, "Hybrid approach based on ga and pso for parameter estimation of a full power quality disturbance parameterized model," *IEEE Trans. Ind. Inf.*, vol. 14, no. 3, pp. 1016–1028, 2018.

[18] F. B. Costa, A. H. P. Sobrinho, M. Ansaldi, and M. A. D. Almeida, "The effects of the mother wavelet for transmission line fault detection and classification," in *Proceedings of the 2011 3rd International Youth Conference on Energetics (IYCE)*, July 2011, pp. 1–6.

[19] W. Gao and J. Ning, "Wavelet-based disturbance analysis for power system wide-area monitoring," *IEEE Trans. Smart Grid*, vol. 2, no. 1, pp. 121–130, 2011.

[20] D.-I. Kim, T. Y. Chun, S.-H. Yoon, G. Lee, and Y.-J. Shin, "Wavelet-based event detection method using pmu data," *IEEE Trans. Smart Grid*, vol. 8, no. 3, pp. 1154–1162, May 2017.

[21] Y.-T. Kao and E. Zahara, "A hybrid genetic algorithm and particle swarm optimization for multimodal functions," *Appl. Soft Comput.*, vol. 8, no. 2, pp. 849–857, Mar. 2008.

[22] R. C. Eberhart and Y. Shi, "Comparison between genetic algorithms and particle swarm optimization." Springer, 1998, pp. 611–616.

[23] T. S. Abdelgayed, W. G. Morsi, and T. S. Sidhu, "Fault detection and classification based on co-training of semisupervised machine learning," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1595–1605, 2018.

[24] M. Mishra, R. Sahu, D. Ray, S. Swarup, and P. K. Rout, "Study of performance of pattern recognition techniques based on wavelet features for Islanding detection in distributed generation," in *Power Electronics, Intelligent Control and Energy Systems (ICPEICES), IEEE International Conference on*. IEEE, 2016, pp. 1–6.

[25] R. Thangaraj, M. Pant, A. Abraham, and P. Bouvry, "Particle swarm optimization: hybridization perspectives and experimental illustrations," *Applied Mathematics and Computation*, vol. 217, no. 12, pp. 5208–5226, 2011.

[26] C. S. Burrus, R. A. Gopinath, and H. Guo, *Introduction to wavelets and wavelet transforms: a primer*. Upper Saddle River, N.J: Prentice Hall, 1998.

[27] [Online]. Available: https://www.naspi.org/work-group-meetings

[28] https://www.dropbox.com/s/wl1aex2s6atmcn1/Data_sheet_microgrid.pdf?dl=0,.