

# Dynamic Event Detection Using a Distributed Feature Selection based Machine Learning Approach in a Self Healing Microgrid

Miftah Al Karim, Jonathan Currie, Tek-Tjing Lie

**Abstract**—A self-healing function is an attractive feature in any modern microgrid (MG). Once a fault occurs it is imperative for a grid to monitor its status, take action based on the level of severity, and after the contingency has been cleared, restore the system. With an increasing number of microgrids and distributed generation stations, deploying a centralized control is no longer a cost-effective option, and therefore distributed control is a likely solution. In an interconnected network it is important to detect the underlying events taking place in each of the distributed stations, otherwise operational decisions become non-coherent. This paper proposes a novel, feature selection-based distributed machine learning approach to detect the dynamic signatures of different power system events. The purpose is to facilitate a post-fault decision-making process in order to restore a stand-alone microgrid without the intervention of a central station. The proposed method detects meaningful features from the generator data and then applies a multi-class classification algorithm to the feature data. Each class represents one dynamic event taking place. The methodology is demonstrated in an interconnected two-area-based microgrid with multiple types of energy generation schemes.

**Index Terms**—Self-Healing, distributed control, feature selection, machine learning

## I. INTRODUCTION

The self-healing feature in a microgrid is a part of the control and operation scheme that introduces a degree of automation [1]. The advent of Phasor Measurement Units (PMU) has improved Supervisory Control and Data Acquisition (SCADA) in recent years. It has also enabled new possibilities in the field of protection and stability analysis, that eventually lead towards self-healing [2]. Through a framework, autonomous systems can be deployed in an interconnected network to maintain system reliability [3]. However, an end-to-end system configuration, which can deploy self-healing mechanisms, may require multiple layers of decision-making processes. These can be represented as a hierarchical, three-tier structure. The bottom layer is placed near each measurement device close to the physical system. The middle layer is intelligence at the substation level and the top layer is the centralized command and control operations undertaken by engineers [4]. The deployment of an immediate control scheme in the bottom layer is considered a well-established strategy to address and solve contingencies locally, without any intervention from the central station. The sole purpose of such an action is to reduce costs and increase speed of operation. The goals of reducing costs and resources have motivated recent authors to explore local decision-making methods [5]–[9]. In [5] a novel sectionalized self-healing approach was

introduced, featuring a sectionalized grid system to mitigate energy distribution problems. The objective set by [5] is to guarantee supply and demand balance in each subsection, either by adjusting the power output from the dispatchable sources or by shedding loads. The article proposed a rolling-horizon optimization method in order to schedule the outputs of the dispatchable distributed generators (DGs). The proposed method has significant merit; however, the article assumes that the self-healing process does not give rise to any dynamic stability issues, which often is not the case [10]. On the contrary, power system stability relies heavily on the dynamic behaviour of the generators. In [6] the authors argue that system instability can be assumed by observing the dynamic behaviour of the generators. In other words, the development of an automated algorithm should rely on the idea of low-level machine control.

Recent literature also suggests that local or distributed control is quite dependent on the identification of dynamic events under different contingencies. An appropriate detection of power system events can lead to smart restoration strategies [1], [2], [6], [11]. This study integrates the concepts of both [5] and [6] and introduces a novel algorithm that detects dynamic events from the distributed generator data in a sectionalized microgrid. The dynamic events are associated with the self-healing process of the microgrid used in this study. The algorithm interprets the dynamic events and decomposes those into user-specified regions to facilitate decision making in the context of restoring an unstable power system. The proposed algorithm has the capacity to detect patterns in the dynamic data and distinguish the data based on the underlying events. Once the underlying event in terms of the affected generator is recognized, the algorithm can take local decisions on each of the generating stations and restore the system after a major fault. The algorithm is installed in each of the generating stations, which are independent of each other, but can make a coherent decision in a post-fault contingency. In the proposed system, two types of non-dispatchable energy sources is introduced; wind and solar. The proposed method is based on a machine learning algorithm which is a modified ensemble of bagged decision trees with an added *Boosting* mechanism. The algorithm uses the generator data collected from the aforementioned grid. For a better prediction, the generator data is augmented using a simplified feature selection process. The purpose of the feature selection process is to prepare the data set for increasing the accuracy of prediction [12]–[14]. However, some of the major concerns regarding feature

selection are high sensitivity to tuning and added redundancy in the algorithmic steps [15]. For these reasons, additional algorithms for improving relevance and removing redundancy go hand in hand with any feature selection process [12]. Solving such challenges would require more resources and would be a time-consuming process to enable data preparation; although this is very important it is unnecessary to address it in a near-real-time application. This study considers an alternative approach, avoiding feature selection in real time by implementing a pre-processed set of input features [14]. The relevance and effectiveness of these features are discussed in the later sections. To develop, train and test the proposed system a two-area-based microgrid was prepared in Matlab-Simulink. The overall contribution of this study can be summarized as:

- 1) Developing a feature selection-based machine learning algorithm that detects dynamic events in a stand-alone and self-healing microgrid.
- 2) Exploring the potential of feature-based augmented datasets, to enhance the predictability of a multiclass classifier algorithm for dynamic time series data.

## II. SYSTEM UNDER CONSIDERATION

The microgrid chosen for this study, as shown in Figure-1, is a two-area system that operates in stand-alone mode. It is a multi-machine system suitable for stability analysis. The purpose of selecting a multi-area microgrid is to separate the system into sections in the event of a fault and apply distributed control. To facilitate such distributed control, each area is equipped with both dispatchable and non-dispatchable distributed generators (DGs). For dispatchable generation three diesel power plants and one hydro power plant based on synchronous generators are considered. The unit local controller in each DG adopts real power frequency and reactive power voltage droop control. For non-dispatchable energy generation, a wind power plant based on an induction generator is considered. Two types of loads are considered; controllable and uncontrollable. Each of the power plants is connected through a medium voltage transmission line of 25kV. The hydro power plant was chosen to cater for the base loads while the diesel power plants address the time-varying loads. The scheduling of the diesel generators was assumed to be based on the availability of renewable energy as well as the demand. The asynchronous generator in the wind power plant has a shunt compensator connected to its node. During high wind power penetration a large inductive load is disconnected from the microgrid. This event triggers a system-wide rotor angle instability due to the generators swinging against each other [16]. The wind turbine is operated at a 94.87% capacitive power factor. Figure-2 shows one illustrative, arbitrarily chosen instance of a 3-second window of post-disturbance rotor angle instability [8], [17], [18]. A 3-second window contains 150 samples. Due to the rotor angle instability the transmission line voltage fluctuates. The proposed algorithm eliminates the rotor angle instability by understanding the dynamic signatures in the generator data and stabilizing the voltage fluctuation in each location,  $\frac{dE_G}{dt} \rightarrow 0$ ;

$E_G$  = Generator terminal voltage. The microgrid used is similar to that in the studies carried out in [7], where a post-fault restorative control scheme is discussed.

Like [5] the system is designed to have both a normal operating mode and a self-healing mode. The self-healing mode consists of series compensators (SCs) in both areas near the synchronous generators. The post-fault contingency is considered as a first swing stability problem, which can be damped using a linear continuous control. From each distributed controller's point of view the grid is a single machine infinite bus system; thus, the active and reactive power transmitted through the transmission line can be modelled as  $P_i = \frac{V^2 \sin \delta}{(1-k)X}$  and  $Q_i = \frac{2V^2}{X} \frac{k}{(1-k)^2} (1 - \cos \delta)$ . Here  $i$  is the distributed generator,  $X$  is the line inductance and  $k = \frac{X_c}{X}$ ;  $X_c$  = series capacitance. Once the proposed algorithm detects an instability it operates the SCs and damps the oscillation. For simplicity the compensators are considered to have two modes; *Bypass Mode* for damping the oscillations in a post-fault scenario and *Blocking Mode* for normal operations. Once a post-fault rotor angle instability is identified, the SC closest to the most affected generator is switched to *Bypass Mode*, to suppress the oscillation and stabilize the system. The SC is then returned to *Blocking Mode* after the system has been restored [19].

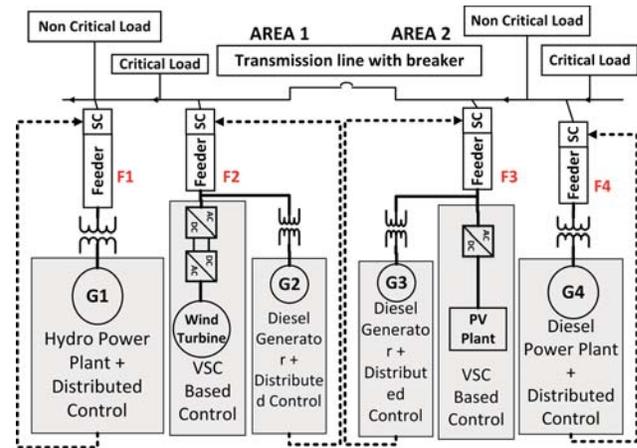


Figure 1. The microgrid model

The model in Figure-1 mostly serves the purpose of generating time series data to train and test the proposed algorithm. The system uses the droop control method for maintaining the frequency and voltage magnitude. The reactive load-sharing capacity of a droop controller largely depends on the feeder impedance. By making the feeder impedance mainly resistive, the aforementioned instability can be introduced. The contingencies are applied in four different locations in close proximity to the four synchronous generators. The proposed distributed method then identifies the generator most affected by any fault and takes relevant action to mitigate it. The algorithm validates the action (decision) by identifying the event that follows. In this study the pre-fault and post-fault events are categorized based on their transient, sub-transient and steady state characteristics.

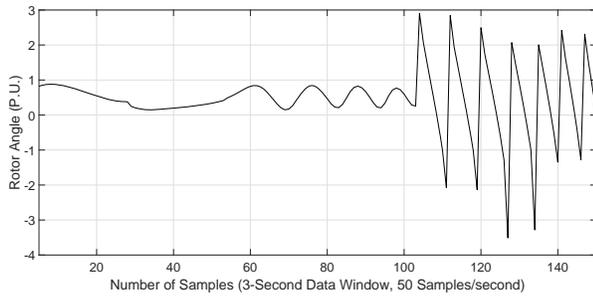


Figure 2. High fluctuation in rotor angle due to a critical fault

### III. PROPOSED METHODOLOGY

The overall workflow of the algorithm in each of the distributed platform, is shown in Figure-3. The algorithm is implemented in a distributed platform at four different generating stations. The workflow shown here, works independently in the distributed locations. It starts with a data preparation stage where a Monte Carlo-based simulation is carried out with the four fault locations near to the wind power plant, solar plant, hydro and diesel power plants. The process is followed by the preparation of a dynamic response database. The database contains the pre-fault, during fault and post-fault dynamic responses of each generator. The feature extraction stage follows, to extract some predefined features and augment the dataset for training machine learning algorithms placed near each of the generators. Once the training process is over the algorithm is evaluated with a new set of randomly chosen datasets.

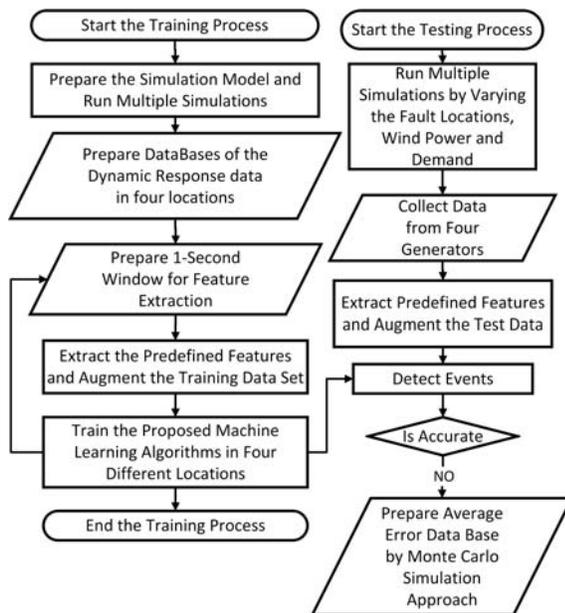


Figure 3. Workflow of The proposed algorithm in each of the four generating stations. This workflow is implemented independently in four locations.

#### A. Data Preparation

The data preparation stage does the data selection and the feature selection shown in Figure-3. In order to develop the

system dynamic response database from the generators, **250** simulations were carried out with randomly selected wind generation and demand for each of the four fault locations as shown in Figure-1 resulting in a total of **1000** different scenarios. For example, the random scenarios relevant to the wind power plant were created with a mean wind speed of  $10m/s$ , and a standard deviation of  $1m/s$ . This study only considers scenarios where rotor angle instability emerges after a major fault.

The feature selection mechanism adopts the concept presented in [20]. The method implies that it should be agnostic to use machine learning for modelling; it should account for a streaming dataset; and it should be scalable to support a large number of events. However, unlike [20], this study implements a fixed feature set to eliminate dependency on any wrapper (different models for different subsets) or filter methodology for feature selection. As the system analysis is largely based on post-fault data, non-linear mapping functions, which are mostly required for load and price signals, can be avoided [15]. Thus a predetermined feature set can be considered. As the time constants of MGs are relatively smaller than those of a high voltage system, generator parameters were carefully considered to select the appropriate features [21]. To produce the generator data a conventional differential algebraic equation-based model was used [22].

This study carried out multiple simulations based on the above principles by randomly varying the wind power and demand while considering four probable fault locations. The overall process of damping the rotor angle instability is mentioned in [7]. However, unlike [7], a supervised control scheme was not implemented; rather the study relied on primary control, secondary control and appropriate series compensation. For simplicity, only cases where oscillation was completely damped were considered as valid candidates for training the machine learning system. Analysis of the undamped scenarios was regarded as out of scope and not undertaken.

To establish a relationship between the generator model and the constant impedance load model, classical energy functions were used [14];

$$M_i \frac{d^2 \delta_i}{dt^2} + D_i \frac{d \delta_i}{dt} = P_i - \sum_{j=1, j \neq i}^m (C_{ij} \sin \delta_{ij} + D_{ij} \cos \delta_{ij}) \quad (1)$$

Where,  $m$  is the number of synchronous generators.  $\delta_i$  is the rotor angle of the  $i$ -th generator.  $C_{ij}$  and  $D_{ij}$  are the function of transfer conductance and susceptance of the reduced network. The per unit inertia constant of each of these generators is  $H = \frac{1}{2} M \omega$ , where  $\omega$  is the synchronous speed.

The dynamic data produced from the generators was selected based on a multi-band power system stabilizer (PSS) model for the hydro power plant and a traditional PSS model for the diesel power plant. The parameters are rotor speed  $\omega$ , rotor angle deviation  $d\delta$ , reactive power generated  $Q_E$  and terminal voltage  $E_G$  [23], [24].

Parameters are represented in per unit. Those generators and consumers not close to the vicinity of the fault are kept constant during the period when a transition between two events is taking place.

## B. System Dynamic Responses

The proposed algorithm implements a simplified feature selection technique for data preparation and applies multiple machine learning algorithms in order to detect underlying events from the time series data. The method is initialized by preparing a database of the system dynamic responses. Four attributes namely: generator rotor speed  $\omega$ , rotor angle deviation  $d\delta$ , reactive power generated  $Q_E$  and terminal voltage  $E_g$  were chosen for preparing the database [6], [16], [18]. The rotor angle deviation was measured against the center of inertia (COI) angle;

$$\delta_{COI} = \frac{\sum_{i=1}^n H_i \delta_i}{\sum_{i=1}^n H_i} \quad (2)$$

$\delta_i$  is the rotor angle and  $H_i$  is the inertia constant of the  $i$ -th generator.

The system dynamic response is characterized by different events taking place in the proposed two-area-based system. The events are divided into nine different categories. The categories or events are based on the transient, sub-transient and steady state behaviour of the pre-fault, during fault and post-fault system states. Two events are associated with the pre-fault state, one event identifies the fault, three events are associated with the post-fault state with classification errors and the remaining three events are associated with the post-fault state without classification errors. For each of these categories or events, generator data was collected from different power plants and stored as a matrix of time series databases. Considering  $t_n$  is the length of time and  $m$  is the number of generators, the attributes of the classification algorithm therefore become  $m$  data points ranging from  $(t_n, E_1, Q_1, \omega_1, d\delta_1)$ ,  $(t_n, E_2, Q_2, \omega_2, d\delta_2)$  to  $(t_n, E_m, Q_m, \omega_m, d\delta_m)$ . The duration of the fault was kept uniform. It was just long enough to introduce a rotor angle instability in the system. However, the length of post-contingent scenarios are not uniform; rather they were chosen based on the time required to achieve system wide stability. This measure was taken to monitor whether a loss of synchronism appears before initiating the next event.

A power system is often interrupted by uncertain events such as three-phase faults. Once the fault is cleared, the system has to be restored and operated at its optimal target configuration. The actual method of system restoration is a well-established research domain that consists of expert systems, mathematical programming and heuristics, as well as soft computing e.g., [11], [25]–[28]. The primary focus of this study is not to analyze different restoration mechanisms, but rather to accurately detect the underlying events and properly categorize them using an automated system. This study assumes a soft computing-based (machine learning-based) system restoration mechanism as discussed in later sections.

The nine events are briefly explained below. Figure-4 shows six time-varying dynamic events out of those nine events.

- 1) Starting the generators: In this state the transient and sub-transient phases of the generators during start-up are addressed.

- 2) Stable operating point after start up: This is the pre-fault stable operating point. Once the system reaches this state the experiment is considered ready for the introduction of faults.
- 3) Introduction of the fault: A large reactive power mismatch is caused by disconnecting a large inductive load. This state lasts long enough so that a rotor angle instability, and in turn, a high amount of power swing, can be introduced.
- 4) Fault clearance and post-fault transient state: In this state the fault is cleared and post-fault transience is observed. The SC is kept in bypass mode (damping mode) in the transmission line.
- 5) Post-fault stable operating point: The steady-state operating point after clearance of the fault.
- 6) Transition towards the initial stable operating point: Once the post-fault system becomes stable the operating conditions are reverted back to the normal pre-fault operating condition. This event represents the transition from the post-fault stable operating point towards the initial restoration states. This state is similar to the sub-transient state of the machine initialization phase.
- 7) Restoration of the initial stable operating point: The stable operating point equivalent to event-2.
- 8) Fault clearance and post-fault transient state: In this state the fault is cleared and post-fault transience is observed. The SC is kept in blocking mode (undamped) in the transmission line. It is a post-fault scenario with misclassification.
- 9) Transition towards the post-fault stable operating point to reach the post-fault stable state or event-5. The SC is kept in blocking mode (undamped) in the transmission line. It is the other post-fault scenario, with misclassification. Event-7 follows both the *event-4* (properly classified decisions) and event-8 (misclassified decisions).

The duration of each of the events may vary depending on the operating conditions of the microgrid. In general, most of the events are longer than one second. During any transition state the later parts become relatively more stable. Thus identification of that half becomes comparatively less challenging due to the reduced variation in data, and also becomes less effective for identifying features. Therefore, to reduce the computational burden, only the attributes exhibiting certain variation in data that exceeds a pre-defined threshold are filtered and sent for feature selection. The rotor speed deviation  $d\omega$  is chosen for preparing this threshold. The crossing of the threshold value is observed for a period of 60ms. In other words there are three consecutive data points  $\Delta\omega_i = \omega_i - \omega_{i-1}; (i \geq 2)$ .

## C. Selected Features

Harnessing information from magnitude and frequency of a signal, has been a well established technique in understanding power quality. Therefore, different signal processing techniques were introduced to understand power quality issues in numerous studies [29]. Scientific analysis in other disciplines such as *Electroencephalography*, signal processing is often

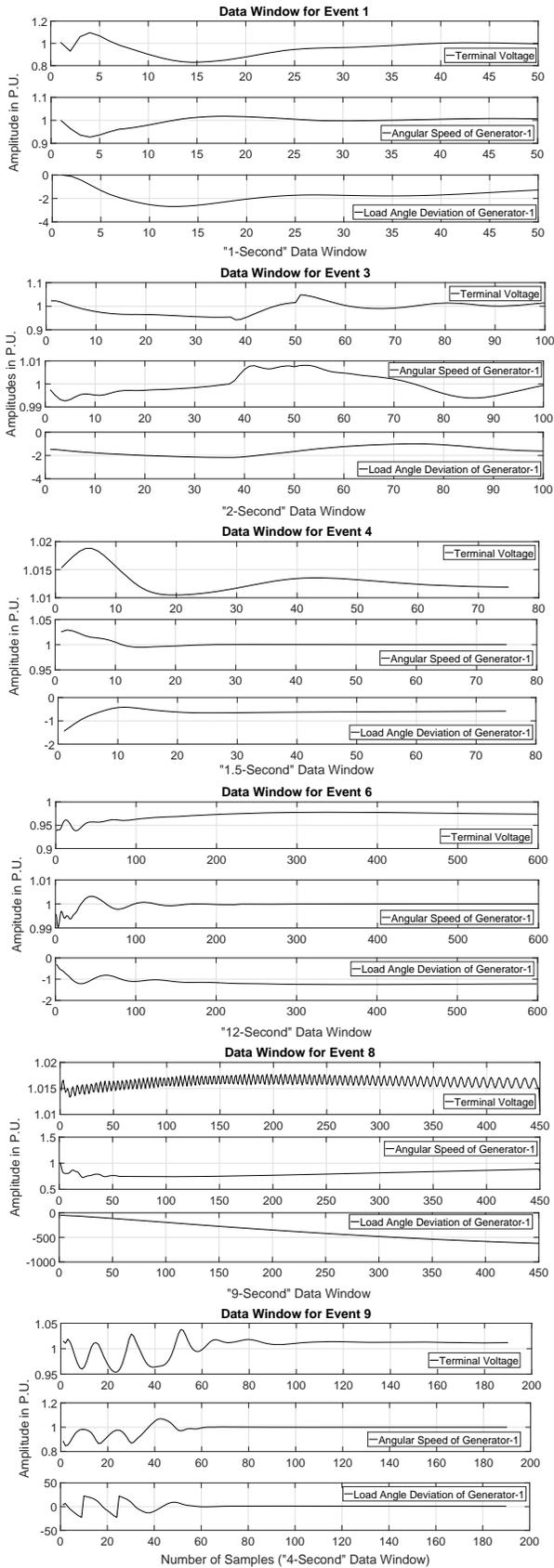


Figure 4. Generator data under different dynamic events. The horizontal axis shows total number of samples (50 samples/second)

implemented to detect events from analyzing the shape of action potential measured in millivolts [30]. The key argument observed in those studies is that, signal shape contains valuable information. This study is also motivated by a similar argument.

After the attribute had been selected, a layer (in the form of attribute columns) of features was added, in order to augment the data to improve predictive capabilities [31], [32]. All the features are normalized to make the overall process generic for different wind power sources and demands. The augmentation is carried out based on three selected engineered features. Based on the shape of a time-series data, these features have the potential to understand an underlying event. The features are:

- 1) Prominence of local maxima.
- 2) The width of the local maxima
- 3) Available frequencies in the time series data

From each of these features, in the next stage, a factor is developed, in order to represent those features as attributes. The attributes are then used to augment the training dataset, to train the machine learning algorithm. The factors, obtained from these features are scalar quantities. The factor derived from the prominence of local maxima is mentioned as prominence factor. The factor derived from the width of the maxima is called width factor. And finally, the factor derived from available frequencies is called the frequency factor in this study. The development of those factors have been further discussed later. The first feature is based on understanding local maxima. It is important in this study, because, the characteristics of local maxima in a data set can carry valuable information [33], [34]. Time series data is composed of different frequencies and amplitudes. The characteristics of those amplitude peaks varies, on the basis of the event taking place at a certain period [35]. Dynamic phasor data obtained from an MG exhibit similar amplitude and frequency variations. This is the primary motivation for selecting the 1<sup>st</sup> and the 2<sup>nd</sup> feature. To select a range in the subject signal, a 1-second window was prepared. The window slides 20ms in each iteration, meaning that after every 20ms one data point was recorded. During the process of calculating features, this 1-second window slides 1-data sample on its right and leaves a data sample on its left, always keeping 50-samples in between. A 1-second window is large enough to retain the features, and small enough not to be corrupter in the transition phases between two events. Further justification for selecting a 1-second is provided later.

As the initial selected feature in the data set, a topographical prominence was chosen. The method followed here is based on the work presented in [36]. The 1-second time series window is first normalized. Then the lowest contour line circulating a local maximum point is detected and the height of that point is measured in terms of that contour line. The second engineered feature is the length of the contour lines. It signifies the width of individual peaks.

Figure-5 shows the method for finding prominent local maxima and width of local maxima. For a better visualization, a 2-second window of the terminal voltage of the generator in hydro power plant (generator-1) has been chosen. However, while

preparing the actual algorithm, always a *1-second* window had been used. Figure-5 represents the prominent local maxima in volts per unit. The width of local maxima has been calculated from the nearest lowest point that encompasses the maxima, shown in the horizontal axis. In the figure, it is observable that local maxima can be of different characteristics. These characteristics can be harnessed as valuable information. The overall process applies Getis and Franklin’s (G&F) variation of Ripley’s K-function rather than a height threshold. The aim is to determine the number of localizations,  $n$ , in the *1-second* data window or the region of interest (ROI). If  $K(r)_j$  is the Ripley’s  $j$ -th K-function and  $\delta_{ij}$  is the Euclidean distance between localization  $\mathbf{i}$  and localization  $\mathbf{j}$ , then;

$$K(r)_j = A \times \frac{\sum_{j=1}^n \delta_{ij}}{n} \quad (3)$$

Here,  $A$  is the area of the ROI and  $r$  is the radius around  $\frac{\sum_{j=1}^n \delta_{ij}}{n}$ , keeping the localization  $\mathbf{j}$  in the center. It means  $K(r)_j$  represents the local signal amplitudes in relation to the average signal amplitude over the whole ROI, which in this case was considered to be the of the peak prominence. The K-function is normalized across its variance and L-function is developed. The diameter  $2 \times r$  signifies the width of the localization  $\mathbf{j}$ , thus the width of local maxima. The L-function and the width of a local maxima has been further processed to develop prominence factor and width factor. The process is explained later with examples.

$$L(r)_j = \sqrt{\frac{K(r)_j}{\pi}} \quad (4)$$

One of the main ways of understanding power system oscillations is frequency spectrum analysis. Due to the spatio-temporal nature of power system dynamics, frequency spectra can develop valuable features [37], which can be helpful not only for understanding inter-area or intra-area oscillations but also for analyzing voltage flicker, as has been done in past research [38]–[40]. The third chosen feature is therefore, derived from the available frequencies in the *1-second* data window. A discrete Fourier transformation (DFT) is used for this feature;

$$x[n] = \sum_{k=0}^{N-1} \frac{1}{N} \tilde{X}[k] e^{-jk(2\pi/N)/n}, K = 0, 1, \dots, N-1 \quad (5)$$

Here,  $\tilde{X}[k]$  is the amplitude and  $x[n]$  is the linear combination of the complex exponentials with that amplitude. DFT has often been used in the field of harmonics and power quality analysis [41]–[43]. In this study, a frequency spectrum of the first  $20Hz$  was chosen to analyze the voltage fluctuation. To convert this information into a feature, sum of the absolute amplitudes,  $|x[n]|$  in Equation-5, where  $n = 1, 2, 3, \dots, 20$  was calculated. It is henceforth mentioned as the frequency factor. The frequency factor is derived from the amplitudes, it has no unit.

$$FrequencyFactor = \sum_{n=1}^{20} |x[n]| \quad (6)$$

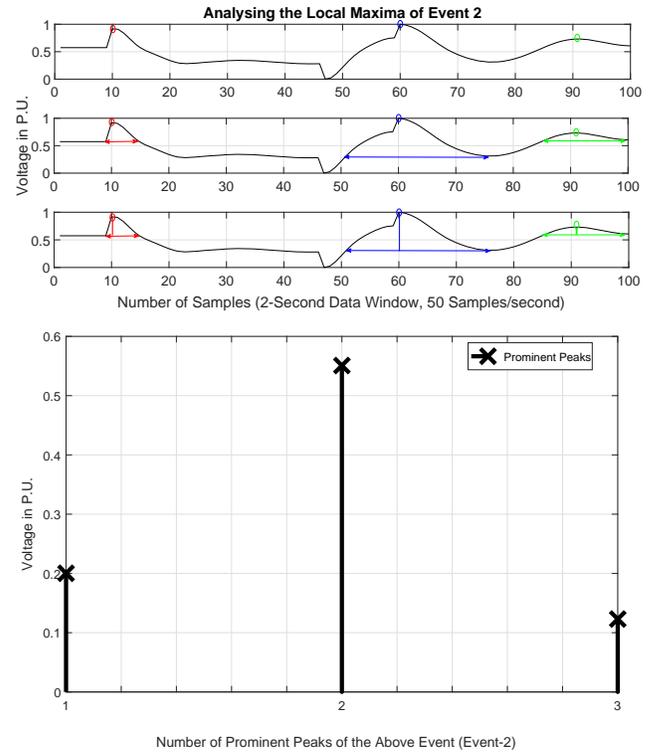


Figure 5. (a) Prominent peaks and their widths inside a predefined data window of the terminal voltage of generator-1. The prominence is presented in volts P.U. (b) The top three peaks, that has been observed in the above data window. The horizontal axis in the bottom sub-figure quantifies the top identified peaks in this example (which is three, in this data sample).

The commercial availability of phasor measurement units has made it easier to store power system attributes at a rate of 50 samples/second or higher. Here, a sampling rate of 50 samples/second was chosen, and thus, based on the “Nyquist–Shannon sampling theorem” the DFT could be carried out up to a frequency of  $25Hz$ . Therefore, *1-second* is sufficient to capture the frequency spectrum in one event. However, in the proposed microgrid it was observed that any frequency more than  $20Hz$  had little to no contribution in calculating the frequency factor. Therefore, the proposed frequency feature is calculated based on frequencies up to  $20Hz$ . Based on the above argument, as mentioned in earlier sections, a window size of *1-second* was chosen. Another motivation for selecting a *1-second* window length was to retain enough distinction between different events. In the proposed microgrid, some of the observed events are no more than two seconds long. Therefore, the frequency features in the transition windows (from unstable towards stable responses), are more likely to be corrupted. The top sub-figure (6a) of Figure-6 shows a comparison between three events under different window sizes. Overall six window lengths (1-6 seconds) have been chosen for this part of the analysis. For each of the window lengths, frequency factors have been calculated for three different events (event-1,3,6). From the figure it is observable that, with different lengths, the frequency factors vary significantly for the same events. The larger the window size, the more corrupted the events become. For example,

being a relatively stable event, event-6 should mostly exhibit a lower value of frequency factors. However, as the window size increases, in the transition phase of event-6, the frequency factor gets corrupted.

The bottom sub-figure (6b) shows a transition window between two events. From the figures it is quite comprehensible that if the data window length increases beyond a certain point, the possibility of identifying distinct features decreases, because characteristics from different events merge.

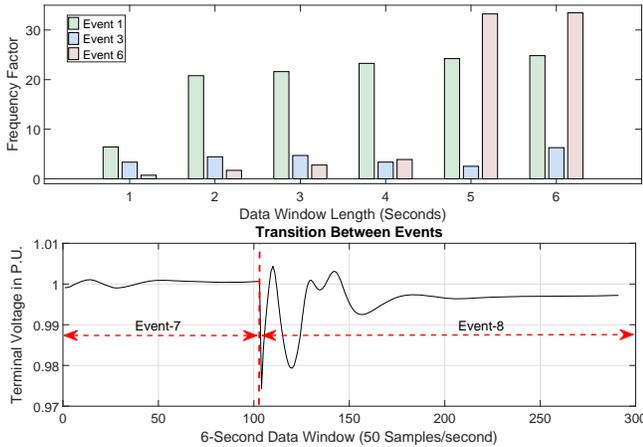


Figure 6. (a) Frequency factor at different window lengths. The frequency factor is represented as the sum of magnitudes, it has no unit. (b) Transition Between Two Events (Event-7 and Event-8). If the data window size increases, the transitional region can merge characteristics from both the events and corrupt their features.

To provide justification for those features, a randomly selected sample data set from three different events was prepared as shown in Figure-7. The chosen attribute was the terminal voltage of generator 1. The figure consists of the samples collected from three events *Event1*, *Event3* and *Event9*. It can be seen that different events, despite having similar overlapping magnitudes, exhibit different characteristics in their signal peaks, widths (Figure-7(a)) and amplitudes of the available discrete frequencies (Figure-7(b)). This justification is reflected in Table-I, where the features, summarized as factors, are shown with respect to different events.

For a better understanding of  $K(r)$  and  $L(r)$  parameters, the aforementioned sample data set as shown in Figure-7 was used. To show the calculation process for peak prominence and width, the *Event1* marked with a black line was considered. Firstly, from any local minimum  $i$  to the next local minimum  $i+1$  a total of  $n = 40$  samples were selected. The distance **26.002** observed in the  $X$ -axis of the data window, between  $i$  and  $i + 1$  is considered as the diameter of the peak. Area  $A$ , **530.92** is then calculated based on this diameter. The average height  $\frac{\sum_{j=1}^n \delta_j}{n}$  of the  $i$ -th peak is calculated by summing the height of each sample  $0.82, 0.83, \dots 1.044, \dots 0.82$  and dividing it by  $n = 40$ , which is **0.4765**. Once the area and average height are found,  $K(r) = \mathbf{253.90}$  and  $L(r) = \mathbf{8.99}$  are calculated using the above-mentioned formula. Each data window can produce multiple  $L(r)$  values. From those multiple  $L(r)$  values the factor representing prominence (prominence factor) is calculated by taking an mathematical mean of all

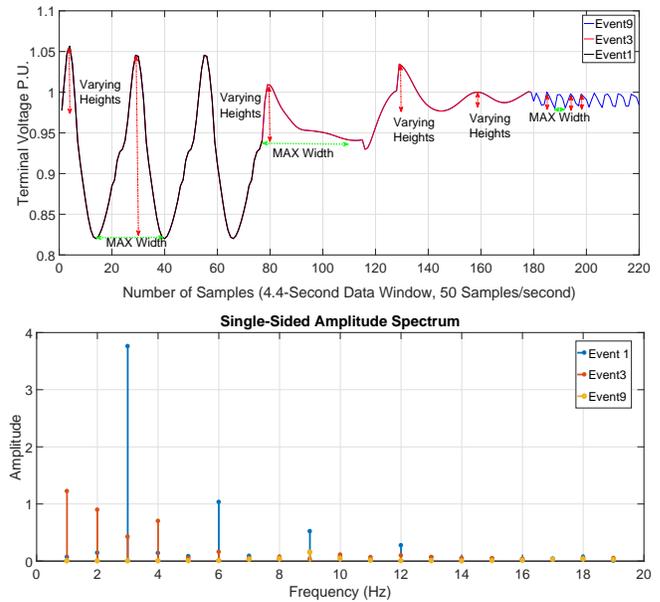


Figure 7. (a) Sample data, representing three different events. These samples are used to show the different features available in those events. (b) Amplitudes obtained from a data window, after carrying out discrete fourier transformation (Eqn-5).

the available  $L(r)$  values in that window. Width of the local maxima is calculated from the diameter **26.002**. To prepare the width factor, maximum width available in a window was selected. A similar data set shown in Figure-7 had also been used to calculate the frequency factors by adding all the amplitudes obtained from DFT of the first  $20Hz$  frequencies in one data window. This process was used to calculate the frequency factors. The feature value with the first  $25Hz$  frequencies is **6.8567**, which is almost equal to the factors observed from the first  $20Hz$  frequencies.

The necessity of preparing different factors, from the features, is quite intuitive. As mentioned earlier, while extracting the features, each data window is sampled with 50 data points per second, representing one measurement every  $20ms$ . However, to train a machine learning algorithm, so that it can identify the underlying event, each of the windows has to be represented as a single row of attributes at any given time. The transition of the data window can thus be considered as a transition of events either from event  $i$  to event  $i$  or from event  $i$  to event  $j$ . From a matrix point of view, the transition of windows means, the transition from one row to the next. In other words, each data window is a row in the training data matrix and the factors are individual attributes in that row. Therefore, an additional action to prepare factors had been taken. This action of preparation of the factors, was carried out in three different ways. For the first factor, a mean value of all the available peak prominences in a data window was selected. This is mentioned as the *prominence factor*. For the second factor, the maximum width among the local maxima was chosen. This feature is mentioned as the *width factor* and for the third feature, summation of amplitudes obtained from the DFT of the first twenty frequencies was considered, known as the *frequency factor*. Thus a range of

50 data points is converted into one data point identifying the feature characteristics of that particular window in a single row of data. Table-I was prepared using the above mentioned factors from the sample dataset.

The intention of the feature selection and data augmentation is to provide additional information to the classification algorithm. In this way significant differences can be seen between two time series data sets that have a high degree of similarity. For example parts of *Event-3*, *Event-6* and *Event-9* are shown in Figure-8. Along with that, the parts of comparatively more stable events, *Event-2*, *Event-5* and *Event-7* are also shown. It can be observed from the figure that, the similarity between these factions is very high for proper classification to be carried out by using the traditional classification methods. Besides, with varying wind power, solar power and demand these similar events overlap each other, which makes it difficult for a machine learning algorithm to gain adequate information.

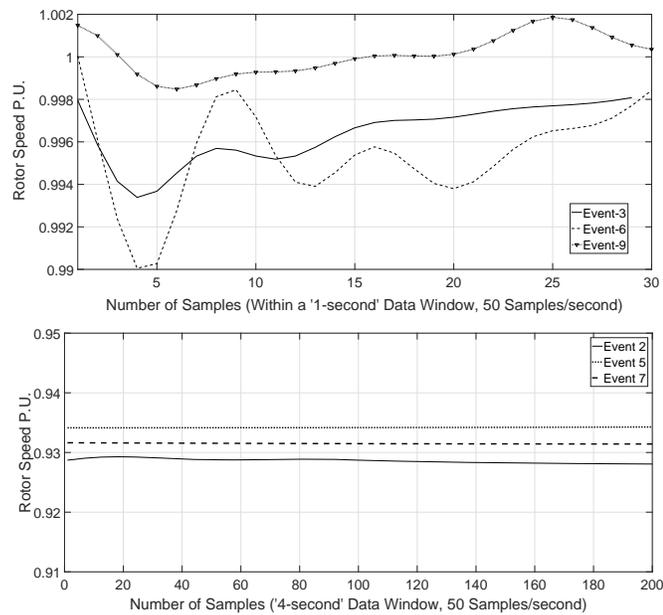


Figure 8. Similarity observed in the time series data among multiple events.

However, adding the feature space to the original data set brings new information to the classifier algorithm with respect to an individual event, as shown in Table-I. The addition of such information helps to reduce classification error and improves the performance of the classification algorithms. This hypothesis is further analyzed, tested and explained in detail in a later section *Testing the Overall Hypothesis*.

#### D. Multiclass Classifier

A multiclass classifier can be an algorithm or a set of algorithms that can predict more than two data classes. To solve such problems, researchers often use ensemble decision trees, sometimes also referred as the random forest. A random forest algorithm uses multiple decision trees and regression techniques, on multiple sub-sets of a training dataset. In this study, once the data preparation stage was over, a multiclass

Table I  
FEATURES (EXTRACTED AS FACTORS) OBSERVED IN THREE SIMILAR EVENTS

Event	Width Factor	Prominence Factor	Frequency Factor
1	26.002	8.93	6.8565
3	33.47	0.94	4.0348
6	47.83	0.042	0.6782
9	4.59	0.001	0.946
2	56.43	0.00035	0.008
5	77.02	0.00001	0.00197
7	109.52	0.00001	0.00111

classifier, which is an ensemble of bagged (bootstrap aggregation) decision trees, was trained as shown in Figure-10. Bootstrapping, divides the training dataset into  $n$ -number of smaller sets with sample replacement, in order to train  $n$ -number of base classifiers. After training, each of those base classifiers votes for a class. The weights of all the votes are then considered in developing one improved composite model, which significantly increases the accuracy of classification [6], [32].

The key advantage of using bootstrap aggregated or bagged classifier is that it enables a linear combination of the function estimates. Based on the different sets of predictor variables and their estimates, the input data is weighted and re-weighted. For example; a function estimate  $\hat{g}_{ens} = \sum_{k=1}^M c_k \hat{g}_k(\cdot)$  is obtained based on the  $k$ -th re-weighted data with a combined linear estimation co-efficient  $c_k$ . This method is helpful in addressing the issue of classification error due to estimation variance and statistical bias, especially in a stochastic scenario [44]. Bagged decision trees are, therefore, not pruned and are longer, which helps the ensemble method to improve accuracy by combining a number of low-biased sub-models. For example, in [40] an ensemble of decision trees was used on PMU-based post-fault rotor angle data, to predict catastrophe in a wide-area power system. The implementation of ensemble trees improved the accuracy of the prediction by more than 20%, compared to a single decision tree.

However, the accuracy of classification or prediction of such an approach, depends heavily on the number of trees. Figure-9 is shown as a proof of concept, where the mean classification errors with different numbers of trees is made. The idea behind this experiment is to show that, increased number of trees can improve classification accuracy. 100 test cases were used for this proof of concept. Each line represents the performance of the proposed algorithm in each of the four power plants. All the algorithms have been fed similar type of data obtained from four different generators. The fault was placed near the hydro power plant and the data of *Event-4* was used as the target attribute. Once the use case was prepared, the classification error of the algorithm was evaluated and the moving average of the error was plotted against the number of trees. The figure indeed portrays that, with an increased number of trees, mean classification error decreases. It was also observed that, if more than 100 decision trees are chosen for the algorithm, accuracy reaches towards a saturation point.

Therefore, based on the above analyses the ensemble was prepared using 100 fully grown trees by splitting the attribute

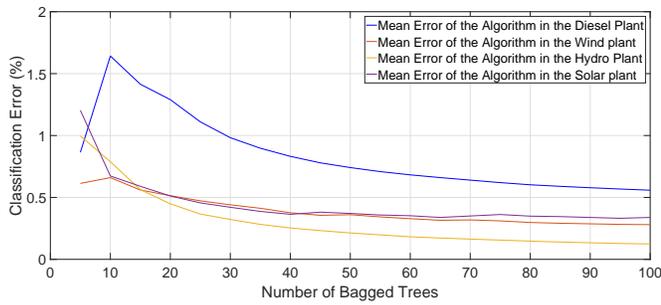


Figure 9. Analysis on classification error vs number of trees. With an increased number of trees, error decreases in all four locations

data into 100 training sets,  $D_1, D_2, \dots, D_{100}$  in order to obtain an improved composite model.  $M_i (1 \leq i \leq 100)$  classifiers vote by predicting a class and the ensemble selects the final class from those votes. The overall technique includes *Bagging* and *Boosting*. The *Boosting* process works in a sequence by considering previously generated classification errors and rectifying those [6], [12], [31], [32]. The classifier is implemented on each of the four synchronous generator data sets. Different contingencies have different durations and thus different dynamic signatures. However, some degree of similarity is present in the dynamic events presented in this study, which affects the process of identifying classes. That is why, unlike [6] and [32], in this study a new scheme of multiple ensemble trees was used to significantly improve the scalability of decision trees.

In this method, using the four attributes of rotor speed  $\omega$ ; rotor angle deviation  $d\delta$ ; reactive power generated  $Q_E$ ; and terminal voltage  $E_g$ , and three features for each of the attributes, three different ensembles of bagged decision trees are trained. Each ensemble then is used to predict the events and an array of predicted events is prepared. The data is then analysed using a simple statistical mode operation and the most frequent prediction in each instance is considered as the final predicted event. This approach has the ability to reduce error while predicting the underlying event. The idea of using the statistical mode is to strengthen the estimation process by incorporating different dynamic responses observed in a synchronous generator. The results are shown in the following section. Figure-10 shows the overall process of preparing the ensemble tree [45]. Hypothesis testing with a limited dataset is discussed in the following section in order to test the proposed method.

### E. Testing the Overall Hypothesis

The overall hypothesis in this study is that different data types, with their inherent features, can significantly differentiate the aforementioned events. The proposed algorithm suffers most during the transition periods from dynamic state to steady state, because the features become less distinct. However, if different data types are used to prepare the features, the accuracy of the ensemble method increases. The idea of using a statistical mode is to filter out the classification errors in

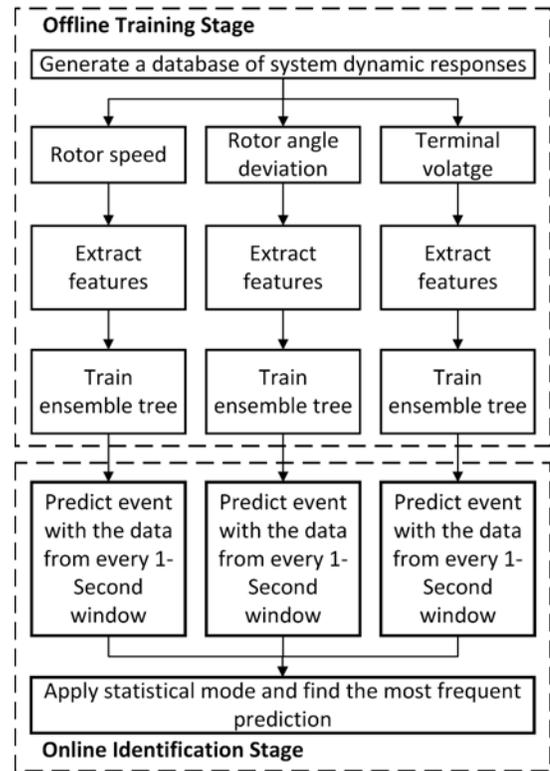


Figure 10. Preparation of the multi-class ensemble trees [6], [45]

one data type by using the results obtained from the other data types.

For example, rotor speed, terminal voltage and reactive power are dynamic in nature and each of them produces similarities between different events at certain durations. However, all of these are not similar at the same data points. Thus if the classifier using rotor speed introduces a classification error, the other classifiers can mitigate that because of their own features. Figure-7 shows the data sample chosen from the terminal voltage of one of the diesel generators for testing the hypothesis.

The hypothesis advocates for feature extraction in order to improve the performance of the classification algorithm. The proposed  $m \times 3$  feature matrix introduced in the earlier section is considered individually irrelevant. The decision tree algorithm recursively tests the attributes and partitions the dataset  $D_t$  containing scalar values into two or more interior nodes  $D_{t1}, D_{t2}$ . Here  $t$  is used to represent a continuous data set representing one window. Each interior node  $D_{ti}$  is defined by a range query such as  $threshold_i \leq D_{ti} \leq threshold_{i+1}$ . However this query range cannot give an appropriate split if the time series data available shares similar values for different classes. Figure-7 is such a time series data set. Where it is clearly observed that, different events have partly similar voltage magnitudes. However, adding the predefined distinct features that represent an individual class helps to eliminate this condition. For clarity, a fictitious example is

provided in Figure-11. The example dataset, plotted in a 2D-plane, has three classes. The intension here is to show that, if feature extraction is used the algorithm reaches the end node (representing only one class) with higher accuracy in fewer steps. On the other hand, without the use of feature extraction, the algorithm reaches the end node with a larger number of steps. The classification error is also high. To demonstrate that, three classes are shown using three colors, yellow, blue and green. The underlying objective is to classify the green class (class-1). In this demonstration two branches from two different trees have been considered. Tree on the left hand side uses shapes (height, width) as features, where the tree on right only uses position of each data point in the X-Y plane. The first data set in any branch is called a root node, and the consequent node obtained from any split is called an interior node (here, only one interior node is shown after a split, as the underlying objective is just to reach any one leaf node that identifies the green class). The interior nodes are also known as non-leaf nodes, where a data set is divided based on a threshold observed in an attribute. The final node that makes a decision or classifies a data set is called a leaf node. As the features are used, the branch of the tree represented on the left hand side reaches to a pure green-class without any error only after two steps (interior node-1, leaf node-1). Despite having some similarities in the position of different data points the features helped gaining more information in each split. While on the right hand side, the branch of the other tree that does not use feature, reaches to an impure green class in three steps. The additional split could not ensure higher accuracy in the later case, rather one misclassified node (yellow dot with red circle) remained. The shape or the feature is increasing information gain.

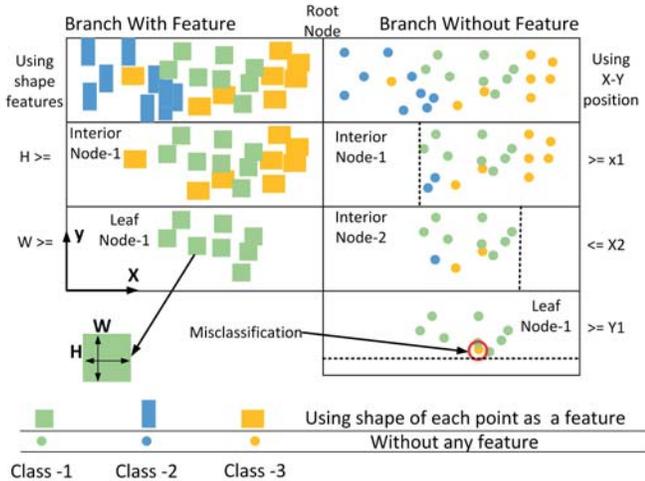


Figure 11. Classification with and without features (fictitious data). Only the decision-paths of the trees that proceed towards root to leaf (Class-1) have been shown here. Other paths have been ignored in this visualization. Misclassification is represented using a hollow red circle.

To further understand the rationale behind the proposed classification approach, a concept called *decision node impurity* is invoked. When the decision tree splits a data set, the node impurity indicates how well the classes are separated. If, in a

particular node, all the data points belong to one class without any classification error, then the node impurity is 0 or the node is considered pure. In Figure-11 the algorithm with feature augmented data reaches a pure leaf-node, while without feature augmentation it reaches to an impure leaf-node.

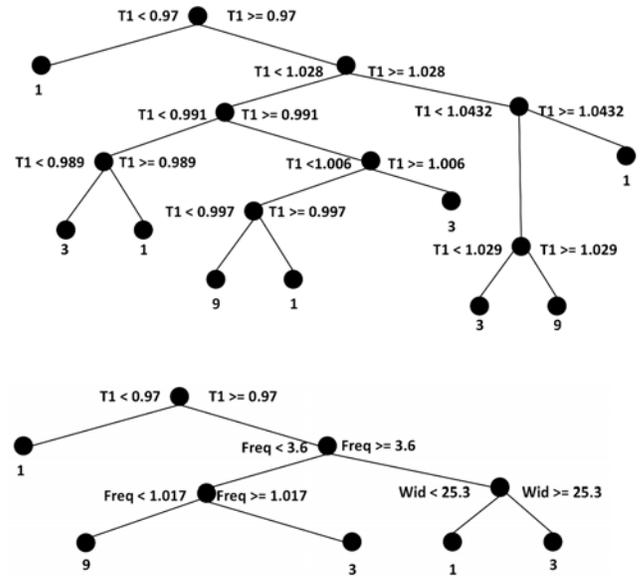


Figure 12. Classification without and with features. The upper tree only uses the time series voltage data  $T1$ , the lower tree uses three features along with the time series data. Here,  $Freq$  = frequency factor and  $Wid$  = width factor as shown in Table-I

Furthermore, in Figure-12 two instances of classification processes with the same data set are presented. This data set is shown in Figure-7. The upper tree is based on only the time series data (terminal voltage of generator-1), and the lower tree is based on the augmented features observed in Table-I. In preparing the upper tree generator terminal voltage (P.U.) is used, denoted as  $T1$ . While, for the lower tree, terminal voltage (P.U.), denoted as  $T1$ , frequency factor, denoted as  $Freq$ , width factor denoted as  $Wid$  and prominence factor have been used. The lower tree that uses features is clearly outperforming the upper tree in number of splits or steps. The upper tree takes higher number of steps. As the upper tree is growing larger than the lower tree it is less efficient. In order to further understand how the tree with feature data is performing better, a determination of goodness of split is carried out.

$$\Delta = I_{Root} - \sum_{j=1}^k \frac{N(v_j)}{N} I(v_j) \quad (7)$$

where  $\Delta$  is the information gain;  $I(Root)$  is the impurity of the root nodes;  $N$  is the number of root nodes;  $k$  is the number of attribute values; and  $N(v_j)$  is the number of interior nodes. The evaluation process measures impurity in order to understand the information gain. The information gain of classifying the first event (*Event1*) was 7.283% for the process without the features and around 26.87% with features. It shows that adding meaningful features reduces impurity. The

Table II  
COMPARISON IN ACCURACIES FOR *Event1*, *Event3* and *Event9* IN PERCENT %

	Generator	Proposed	CART	ANN	K-NN
	Event-1	1	99.8949	89.4529	< 30
2		99.8603	89.7265	< 30	93.5394
3		98.9123	89.3287	< 30	90.6113
4		95.7268	92.4601	< 30	95.4763
Event-3	Generator	Proposed	CART	ANN	K-NN
	1	99.6636	87.3372	< 30	90.0813
	2	99.6765	87.52398	< 30	91.5239
	3	99.6042	88.4577	< 30	92.1306
	4	97.9386	89.6389	< 30	93.9854
Event-9	Generator	Proposed	CART	ANN	K-NN
	1	97.1287	81.0249	< 30	89.8023
	2	97.1146	84.2273	< 30	91.2254
	3	97.5543	79.4611	< 30	93.1175
	4	95.2584	73.7433	< 30	94.0028

comparison is done based on the first three interior nodes. The impurity was then calculated based the equation shown below for each of the root and interior nodes [44], [46], [47]. In the equation,  $N_i$  is the node impurity,  $C_{target}$  is the number of samples representing the target class and  $C_{total}$  is the total number of samples those are available in the interior node.

$$N_i = \frac{C_{total} - C_{target}}{C_{total}} \quad (8)$$

For example, if in a interior node the observed  $C_{target}$  is 50 samples of *event-3* and in the same node  $C_{total}$  is 100 samples of *event-3* and *event-9*, then the node impurity is 50%

#### IV. RESULTS

The proposed method was then examined and compared using a test data base. The test database was prepared by randomly distributing the events. The test set contains  $400 \times 9$  events scattered over the time frame. To test the robustness of the algorithm, a normal random noise was also added to the attribute data. The random noise had a mean of **0** and standard deviation of **0.01** P.U. Table-II shows the accuracy of the algorithm in detecting the dynamic events in each of the four generating stations. Table-II also shows a comparison between the performance of the algorithm and three other traditional machine learning-based methods: a classification and regression tree (CART); an artificial neural network (ANN); and a K-nearest neighbour algorithm. It is observed that a significant improvement can be achieved if features are introduced to the time series data that can add a distinct attribute characterizing each of the underlying events. Classification errors mostly occur during the transition between one event and another event, especially when the system is achieving stability. This is because when a system is near a stable state, features become less prominent and the margin of error increases, as observed in case of event-9 in the Table-II.

The comparison clearly demonstrates the superiority of a feature selection-based method over traditional classification methods. Some of the classical techniques are very accurate and promising. However, as a stand-alone microgrid is a highly sensitive network, a small classification error in a post-fault

scenario can have significant consequences. Therefore, it is desirable to implement an algorithm with improved accuracy.

In Figure-13 the overall accuracy in detecting all the **nine** events of the proposed distributed method is determined based on **four** different fault locations. In total **400** simulations were carried out to test the algorithm; **100** simulations for each of the fault locations. In each simulation the wind power and demand was randomly varied within a predefined boundary to develop different test data. The *black*, *blue*, *green* and *red* lines represent the accuracy of the algorithm in the diesel power plant, the solar power plant, the wind plant and the hydro power plant respectively. While the *Dotted ...*, the *Dashed - -*, the *Dot-Dashed -.-* and the *Solid* lines represent faults near the solar power plant, the diesel power plant, the hydro power plant and the wind power plant respectively. After each simulation, Figure-13 was updated and the running means of the observed accuracies were plotted. Therefore, the convergence observed in accuracy, represents the overall accuracy, once all the test cases have been performed. The data in Figure-13 shows dynamic behavior because after each simulation the mean accuracy of the algorithm is updated. That is why, the more simulation is carried out the more convergence is achieved in mean accuracy.

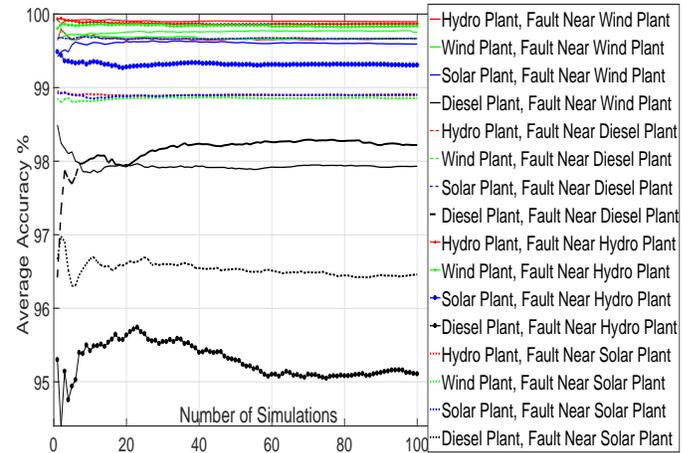


Figure 13. Running mean of accuracy of the algorithm in four locations, observed after a total of 400 ( $4 \times 100$ ) simulations. With four generators and four fault-locations, sixteen combinations have been observed.

The key application of the proposed method is likely to be in the context of identifying the generator affected most by a fault and making valuable decisions in order to clear the impact and restore the grid.

Based on the determination of the underlying event, the distributed controllers operate the SC located closest to the faulty generator. This means the decision is to identify which SC placed near to the subject generators should be in bypass mode. The other SCs will remain in the blocking or non-operational mode. Table-III represents all the available features for a *1-second* data window during the fault period. The fault location was altered from close to generator-1 to generator-4 and four scenarios were created. The algorithms are distributed thus, do not share data between themselves. The application shown in column **6** shows whether an SC should be operated in bypass mode or not in order to damp the oscillation. **Yes**

Table III  
ONLINE DECISION MAKING BASED ON THE FEATURE DATA COLLECTED FROM DIFFERENT LOCATIONS (OBSERVED FROM DIFFERENT GENERATORS)

Observed From	Fault at	Prominence Factors	Width Factors	Frequency Factors	Decision
G1	G4	0.0315	38.878	12.1	No
G1	G3	0.00166	34.72	1.2	No
G1	G2	0.038	37.57	12.4	No
G1	G1	10.64	27.183	49.98	Yes
G2	G4	0.035	21.51	15.2	No
G2	G3	0.00135	22.35	2.7	No
G2	G2	8.66	19.57	37.2	Yes
G2	G1	0.044	21.21	22.3	No
G3	G4	0.038	27.5	8.66	No
G3	G3	7.4	19.88	4.74	Yes
G3	G2	0.002	24.07	5.76	No
G3	G1	0.018	24.3	9.8	No
G4	G4	9.02	26.1	35.46	Yes
G4	G3	0.0002	34.8	9.12	No
G4	G2	0.017	34.75	10.03	No
G4	G1	0.025	35.44	10.82	No

means operate the closest SC and **No** means do not operate that SC. Each algorithm can only control the SC located closest to it. Once the algorithm recognizes the transition from **Event3** towards **Event4** the algorithm switches back the SCs into non-operational mode. This switching of the SCs based on the identification of different events is considered an *online* operation. From Table-III it is observable that, adding different features, can help distinguish event and facilitate decision making. It is because, when a fault occurred near any power plant, the features change accordingly in all the four locations.

Figure-14 shows a simulated event where a decision error results from a classification error. The decision made with a classification error is then compared with a decision without any classification error. In this event the algorithm placed closer to the Generator-1 fails to identify the location of the fault and does not operate its SC but Generator-4 operates the SC near it, thus affecting the direction of power flow.

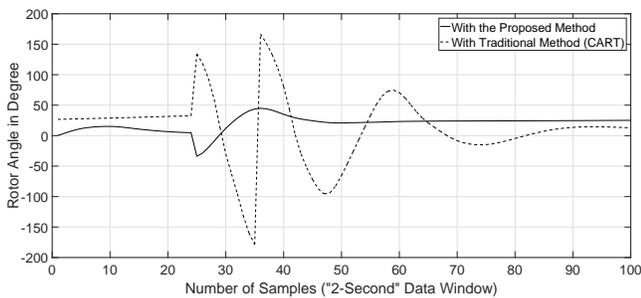


Figure 14. Decision making with and without the proposed algorithm. The proposed algorithm can detect an event faster with higher accuracy

Figure-15 shows a post-fault operation based on the decisions made. This shows how the distributed SCs are operated once the fault location is detected near generator-1. Only the SC close to generator-1 is operated in bypass mode and the oscillation is damped to achieve a post-fault stable state. The switching in SC takes place immediately after the fault is

cleared.

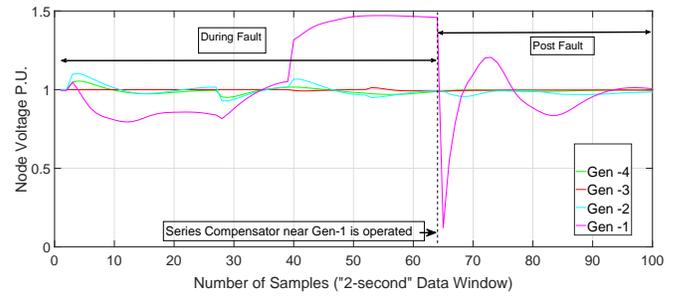


Figure 15. Post fault operation of generator-1 with the nearby SC in bypass mode. The algorithms placed in all four locations are working independently.

## V. CONCLUSION

This study develops and demonstrates a novel algorithm to detect power system dynamic events by implementing a less computationally expensive feature selection method. The performance of the algorithm is clearly superior to the traditional classification approaches. Minimal classification errors were found to occur during the transition periods from one event to another. Once the data window falls into only one category the classification becomes close to 100% accurate.

The performance of the algorithm was tested on an offline basis only. Any misclassification due to time lag or missing data was ignored. Another key consideration with the current algorithm is selecting an appropriate data window. Faults are considered to have a fixed length before being cleared. During each of the short data windows, loads as well as non-dispatchable generation are considered constant. In future studies, dynamicity of the loads, intermittent generation and fault duration will also be addressed.

The detection of events is mostly based on one type of three-phase fault leading to rotor angle instability. More comprehensive analysis needs to be carried out for other types of instability. Furthermore, the decision-making capability can be made more extensive by considering optimal restoration actions for a cost-effective post-fault operation.

Overall, it can be stated that the proposed algorithm has shown promise, despite being less computationally intensive than the traditional methods.

## REFERENCES

- [1] F. Shahnia, R. P. S. Chandrasena, S. Rajakaruna, and A. Ghosh, "Primary control level of parallel distributed energy resources converters in system of multiple interconnected autonomous microgrids within self-healing networks," *IET Generation, Transmission Distribution*, vol. 8, pp. 203–222, February 2014.
- [2] 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 Transactions on Power Delivery*, vol. 32, pp. 294–301, Feb 2017.
- [3] D. Q. Oliveira, A. C. Z. de Souza, A. B. Almeida, M. V. Santos, B. I. L. Lopes, and D. Marujo, "Microgrid management in emergency scenarios for smart electrical energy usage," in *PowerTech, 2015 IEEE Eindhoven*, pp. 1–6, June 2015.

- [4] M. Amin, "A smart self-healing grid: in pursuit of a more reliable and resilient system [in my view]," *IEEE Power and Energy Magazine*, vol. 12, no. 1, pp. 112–110, 2014.
- [5] Z. Wang and J. Wang, "Self-healing resilient distribution systems based on sectionalization into microgrids," *IEEE Transactions on Power Systems*, vol. 30, pp. 3139–3149, Nov 2015.
- [6] T. Guo and J. V. Milanović, "Online identification of power system dynamic signature using pmu measurements and data mining," *IEEE Transactions on Power Systems*, vol. 31, pp. 1760–1768, May 2016.
- [7] M. A. Karim, J. Currie, and T. T. Lie, "A distributed machine learning approach for the secondary voltage control of an islanded micro-grid," in *2016 IEEE Innovative Smart Grid Technologies - Asia (ISGT-Asia)*, pp. 611–616, Nov 2016.
- [8] J. He, Y. W. Li, and F. Blaabjerg, "An enhanced islanding microgrid reactive power, imbalance power, and harmonic power sharing scheme," *IEEE Transactions on Power Electronics*, vol. 30, pp. 3389–3401, June 2015.
- [9] I. Kouveliotis-Lysikatos and N. Hatziaargyriou, "Distributed economic dispatch considering transmission losses," in *2017 IEEE Manchester PowerTech*, pp. 1–6, June 2017.
- [10] M. Amin, "Toward self-healing energy infrastructure systems," *IEEE Computer Applications in Power*, vol. 14, pp. 20–28, Jan 2001.
- [11] T. Nagata and H. Sasaki, "A multi-agent approach to power system restoration," *IEEE Transactions on Power Systems*, vol. 17, pp. 457–462, May 2002.
- [12] C. Zhang, Y. Li, Z. Yu, and F. Tian, "Feature selection of power system transient stability assessment based on random forest and recursive feature elimination," in *2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, pp. 1264–1268, Oct 2016.
- [13] K. Niazi, C. Arora, and S. Surana, "Power system security evaluation using ann: feature selection using divergence," *Electric Power Systems Research*, vol. 69, no. 2, pp. 161–167, 2004.
- [14] J. Geeganage, U. D. Annakkage, T. Weekes, and B. A. Archer, "Application of energy-based power system features for dynamic security assessment," *IEEE Transactions on Power Systems*, vol. 30, pp. 1957–1965, July 2015.
- [15] O. Abedinia, N. Amjadi, and H. Zareipour, "A new feature selection technique for load and price forecast of electrical power systems," *IEEE Transactions on Power Systems*, vol. 32, pp. 62–74, Jan 2017.
- [16] E. Vittal, M. O'Malley, and A. Keane, "Rotor angle stability with high penetrations of wind generation," *IEEE Transactions on Power Systems*, vol. 27, no. 1, pp. 353–362, 2012.
- [17] J. He and Y. W. Li, "Analysis, design, and implementation of virtual impedance for power electronics interfaced distributed generation," *IEEE Transactions on Industry Applications*, vol. 47, pp. 2525–2538, Nov 2011.
- [18] Z. Wang, E. B. Makram, G. K. Venayagamoorthy, and C. Ji, "Dynamic estimation of rotor angle deviation of a generator in multi-machine power systems," *Electric Power Systems Research*, vol. 97, pp. 1–9, 2013.
- [19] F. M. Tehrani, G. Shahgholian, and H. Pourghassem, "Dynamic study and stability analyze of damping cohesfion and reactance in tesc controller connected on optimization smib system," in *2011 IEEE 3rd International Conference on Communication Software and Networks*, pp. 270–274, May 2011.
- [20] M. D. Hausmann and H. Ziekow, "The potential of household specific feature selection for analysing smart home time-series data," in *Smart SysTech 2016; European Conference on Smart Objects, Systems and Technologies*, pp. 1–6, July 2016.
- [21] P. Ferraro, E. Crisostomi, M. Raugi, and F. Milano, "Analysis of the impact of microgrid penetration on power system dynamics," *IEEE Transactions on Power Systems*, vol. PP, no. 99, pp. 1–1, 2016.
- [22] D. J. Hill and I. M. Y. Mareels, "Stability theory for differential/algebraic systems with application to power systems," *IEEE Transactions on Circuits and Systems*, vol. 37, pp. 1416–1423, Nov 1990.
- [23] P. Arunagirinathan, H. A. Abdelsalam, and G. K. Venayagamoorthy, "Remote power system stabilizer tuning using synchrophasor data," in *2014 IEEE Symposium on Computational Intelligence Applications in Smart Grid (CIASG)*, pp. 1–7, Dec 2014.
- [24] R. Grondin, I. Kamwa, G. Trudel, L. Gerin-Lajoie, and J. Taborda, "Modeling and closed-loop validation of a new pss concept, the multi-band pss," in *2003 IEEE Power Engineering Society General Meeting (IEEE Cat. No.03CH37491)*, vol. 3, p. 1809 Vol. 3, July 2003.
- [25] T. Nagata, Y. Tao, H. Sasaki, and H. Fujita, "A multiagent approach to distribution system restoration," in *2003 IEEE Power Engineering Society General Meeting (IEEE Cat. No.03CH37491)*, vol. 2, p. 660 Vol. 2, July 2003.
- [26] L. Sun, C. Zhang, Z. Lin, F. Wen, Y. Xue, M. A. Salam, and S. P. Ang, "Network partitioning strategy for parallel power system restoration," *IET Generation, Transmission Distribution*, vol. 10, no. 8, pp. 1883–1892, 2016.
- [27] D. R. Medina, E. Rappold, O. Sanchez, X. Luo, S. R. R. Rodriguez, D. Wu, and J. N. Jiang, "Fast assessment of frequency response of cold load pickup in power system restoration," *IEEE Transactions on Power Systems*, vol. 31, pp. 3249–3256, July 2016.
- [28] N. Mithulananthan, C. A. Canizares, J. Reeve, and G. J. Rogers, "Comparison of pss, svc, and statcom controllers for damping power system oscillations," *IEEE transactions on power systems*, vol. 18, no. 2, pp. 786–792, 2003.
- [29] M. H. Bollen and I. Gu, *Signal processing of power quality disturbances*, vol. 30. John Wiley & Sons, 2006.
- [30] S. Sanei and J. A. Chambers, *EEG signal processing*. John Wiley & Sons, 2013.
- [31] J. Heaton, "An empirical analysis of feature engineering for predictive modeling," in *SoutheastCon 2016*, pp. 1–6, March 2016.
- [32] P. H. R. N. K. K. C. T. L. N. K. and B. R. M., "Feature engineering on forest cover type data with ensemble of decision trees," in *2015 IEEE International Advance Computing Conference (IACC)*, pp. 1093–1098, June 2015.
- [33] R. Ueda, T. Arai, K. Sakamoto, Y. Jitsukawa, K. Umeda, H. Osumi, T. Kikuchi, and M. Komura, "Real-time decision making with state-value function under uncertainty of state estimation-evaluation with local maxima and discontinuity," in *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, pp. 3464–3469, April 2005.
- [34] I. S. Honkala and T. K. Laihonon, "The probability of undetected error can have several local maxima," *IEEE Transactions on Information Theory*, vol. 45, pp. 2537–2539, Nov 1999.
- [35] E. W. Abel, N. S. Arkidis, and A. Forster, "Inter-scale local maxima-a new technique for wavelet analysis of emg signals," in *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol.20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No.98CH36286)*, vol. 3, pp. 1471–1473 vol.3, Oct 1998.
- [36] J. Griffié, L. Boelen, G. Burn, A. P. Cope, and D. M. Owen, "Topographic prominence as a method for cluster identification in single-molecule localisation data," *Journal of biophotonics*, vol. 8, no. 11–12, pp. 925–934, 2015.
- [37] E. Barocio, B. C. Pal, N. F. Thornhill, and A. R. Messina, "A dynamic mode decomposition framework for global power system oscillation analysis," *IEEE Transactions on Power Systems*, vol. 30, pp. 2902–2912, Nov 2015.
- [38] Y. Y. Hong and L. H. Lee, "Analysis of equivalent 10 hz voltage flicker in power systems," *IEE Proceedings - Generation, Transmission and Distribution*, vol. 146, pp. 447–452, Sep 1999.
- [39] M. V. E. Xavier, W. d. C. Boaventura, and M. D. Noronha, "High performance power quality monitoring system," in *2016 17th International Conference on Harmonics and Quality of Power (ICHQP)*, pp. 974–979, Oct 2016.
- [40] I. Kamwa, S. Samantaray, and G. Joos, "Catastrophe predictors from ensemble decision-tree learning of wide-area severity indices," *IEEE Transactions on Smart Grid*, vol. 1, no. 2, pp. 144–158, 2010.
- [41] M. V. E. Xavier, W. d. C. Boaventura, and M. D. Noronha, "High performance power quality monitoring system," in *2016 17th International Conference on Harmonics and Quality of Power (ICHQP)*, pp. 974–979, Oct 2016.
- [42] Z. Hu, M. Xiong, H. Shang, and A. Deng, "Anti-interference measurement methods of the coupled transmission-line capacitance parameters based on the harmonic components," *IEEE Transactions on Power Delivery*, vol. 31, pp. 2464–2472, Dec 2016.
- [43] G. L. G. de Pinho, G. S. Wojchowski, and C. D. P. Crovato, "Tool for power quality assessment of motor fed by variable speed electric drivers," in *2016 17th International Conference on Harmonics and Quality of Power (ICHQP)*, pp. 109–114, Oct 2016.
- [44] P. Bühlmann, "Bagging, boosting and ensemble methods," in *Handbook of Computational Statistics*, pp. 985–1022, Springer, 2012.
- [45] P. C. Wang, "Packet classification with multiple decision trees," in *2015 21st Asia-Pacific Conference on Communications (APCC)*, pp. 626–631, Oct 2015.
- [46] I. Guyon and A. Elisseeff, "An introduction to feature extraction," *Feature extraction*, pp. 1–25, 2006.
- [47] P.-N. Tan, M. Steinbach, and V. Kumar, "Classification: basic concepts, decision trees, and model evaluation," *Introduction to data mining*, vol. 1, pp. 145–205, 2006.