A New Approach for Machine Learning-Based Fault Detection and Classification in Power Systems

Halil Alper Tokel, Rana Al Halaseh, Gholamreza Alirezaei, and Rudolf Mathar
Institute for Theoretical Information Technology
RWTH Aachen University, D-52056, Aachen, Germany
{tokel,alhalaseh,alirezaei,mathar}@ti.rwth-aachen.de

***Abstract*—The detection and classification of faulty conditions in power systems is a task of crucial importance for a reliable operation. Recently, the use of high-resolution synchronized phasor measurements has been proposed by several researchers for fault detection and classification. Unlike the proposed approaches available in the literature, the central idea in this work is to leverage the delay information of phasor measurement streams to enable a faster recognition of faulty operation. In this work, therefore, we focus on the effect of the communication network delays on the fault detection time, and propose a novel training technique for fault detection and classification which takes delayed measurements into consideration. The performance of the proposed approach is verified using simulated power system data, where artificial neural networks are used for fault detection and classification.**

***Index Terms*—power system, fault detection, fault classification, machine learning, artificial neural network**

I. Introduction

The power generation, transmission, and distribution systems are the key enablers of the modern economy and our sophisticated daily lives. Their reliable operation is not only crucial to prevent economical losses due to outages but also to keep all vital services, such as health care and communication, available night and day. In 2003, around 50 million people in North America were affected by one of the greatest blackouts of the history causing an estimated economical loss between 4 and 10 billion USD to the US economy. The lack of situational awareness on the network was identified as the primary cause of the sequence of events leading to this blackout [1]. This conclusion triggered an unprecedented effort of modernization and development towards a *smart grid* which incorporates sophisticated measurement and communication infrastructures in a wider scale to ensure a real-time monitoring and an online optimization of the power system. Similarly, the increasing integration of renewable energy brought about the requirements of enhanced state-information into the distribution network level. A tangible result of this development is the ongoing deployment of phasor measurement units (PMU) and wide area measurement systems (WAMS).

The high resolution data provided by PMUs open up new possibilities also for the detection and classification of anomalous operating conditions. On the one hand, the synchronized and time-stamped measurements from PMUs make it possible to observe and compare network voltages which accurately reflect the actual system state at the measurement time. On the other hand, the analysis of these data is a challenging task. Therefore, the application of machine learning-based techniques on PMU data has been recently proposed by numerous researchers and continues to attract more attention [2]–[5]. A review of fault detection and classification techniques can be found in [6].

Artificial intelligence-based protection of power systems is not entirely new and already dates back to at least early 1990s, see for example [7] and [8]. However, the advance of high resolution and time-stamped PMU measurements brings up new opportunities for designing and realizing reliable systems based on data analytics. For example, a classification method based on decision trees is proposed in [2] where real PMU data from a network operator is used. Similarly, simulated voltage phasor values from a *PMU-only state estimator* are used to train a classification and regression tree in [3]. Both works show the potential and accuracy of detection and classification techniques based on machine learning. Furthermore, it is shown in [5] that an artificial neural network (ANN) can be reliably used for fault detection on power transmission lines.

One significant conclusion from the analysis of real network data in [2] is that faulty events leave detectable impacts in the signals of multiple PMUs which are located in the neighborhood of the fault position. Therefore, the usage of measurements across the whole grid offers a potential for increasing the accuracy of correct detection and classification as stated in [2]. Having said that, the training methods proposed in [2] and [3] use synchronized measurements at the steady state after the fault occurred. However, the communication delays in WAMS and the processing delays at the phasor data concentrators (PDC) result in different delays in the arrival of measurements from different PMUs in a data management center, the so-called super PDC (SPDC). Furthermore, the failure of one or several PMUs can result in outdated information. With this motivation in mind, we propose a new training method for machine learning-based detection and classification schemes under consideration of communication delays. In this way, the system condition will be trained as faulty as soon as the first measurement, which is recorded after the fault, arrives at the SPDC. In the present paper, hence, we investigate the potential and applicability of this approach to improve the detection and classification time.

The rest of this paper is structured as follows. We start with a brief discussion of WAMS and its operation along with a

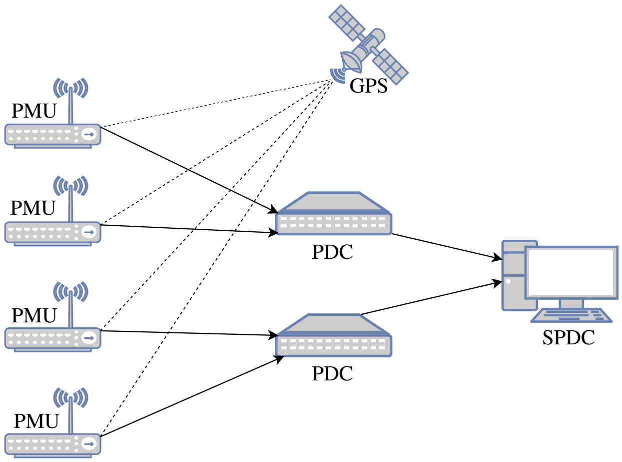

Fig. 1. Hierarchical network architecture of a WAMS. The PMUs send the phasor measurements, that are time-stamped by the GPS signal, to a SPDC over intermediate PDCs [9].

note on the potential of using communication network metrics in situational awareness applications. Next, the fault types in power systems and their detection are briefly introduced, as well as the concepts of ANNs. In Sec. IV, we present the proposed training scheme along with the details of the simulations. Finally, we discuss the results of the simulation study and conclude the paper with final remarks.

II. WAMS Operation and Communication Delays

A WAMS consists of *i)* PMUs, which measure the voltage and current phasor values available at the system nodes where they are installed, *ii)* several PDCs, and *iii)* a SPDC. The *IEEE Standard for Synchrophasor Data Transfer for Power Systems* [9] postulates a hierarchical transmission of sensor data from PMUs to PDCs as shown in Fig. 1, where a preprocessing of the data takes place such as a time alignment and a consistency check [10]. The network may also have multiple layers of PDCs, where the data are hierarchically aggregated and sent to higher layers. The measurement data from a larger part of the network are aggregated in the central unit SPDC to execute energy management functions such as a state estimation.

As further specified in the related standard for data concentrators [11], for each time stamp, the PDCs align and combine the received measurements. In this way, the measurements belonging to a single time stamp are filled into a single data packet and forwarded to the SPDC. In this task, the PDC waits for each time stamp a certain amount of time, referred to as *wait time*, before sending the packet. Hence, the measurements experiencing a larger communication delay than the wait time will not be available at SPDC with this packet. Moreover, the standard discusses and allows the deactivation of time alignment for time-critical applications, in which all measurements are forwarded to the SPDC without a wait time. Therefore, the observation of the system state will consist of measurements with different, but known delays. The state estimation function can also make use of this information by

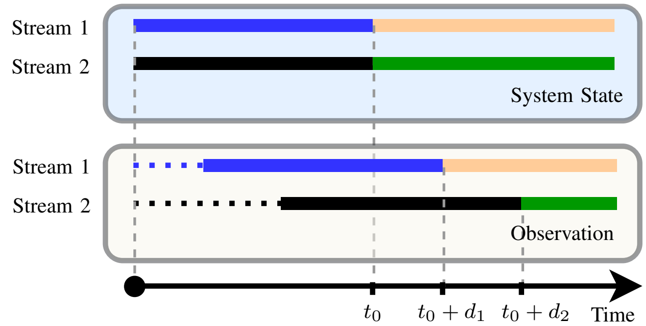

Fig. 2. A possible observation of two PMU streams, which originate from different PMUs. A fault causes detectable changes (illustrated by a change of color) in signal patterns at time t_0, whereas the changes are observed at time $t_0 + d_1$ and $t_0 + d_2$ at SPDC for Stream 1 and 2, respectively.

assigning weights to different measurements in the estimation process based on their delay, for example with a weighted least squares estimator [12]. Therefore, the output of the state estimator inherently contains information about the network delays. An illustration of the delayed arrival of measurements is provided in Fig. 2 for two measurement streams which originate from different PMUs. Note that the pattern changes in the signals, which are illustrated by a change of color at t_0 due to a fault, are observed at the SPDC with delays of d_1 and d_2, respectively. In addition to the available measurement values, the information of the delays is also available due to the time stamps. The central idea of this work is to investigate the potential of leveraging this information in order to detect the system disturbances in a faster and reliable manner.

In fact, the potential use of time information included in measurements has recently been proposed by the Data Mining Initiative of Electric Power Research Institute (EPRI) [13]. In its published data analytics case "Sequence of Outage Events Replay" the description of the case states [14]:

> *"In response to a fault or a series of faults on an electrical distribution circuit, tens if not hundreds of distribution devices report their status to system operators for the purpose of managing the outage. Timing information that can be used to chronicle the system response is present in the data, but it is generally not leveraged - resulting in missed opportunities to analyze the dynamic response of the distribution system."*

The core motivation of this work is, therefore, closely linked to the idea which is identified and put forward by EPRI.

III. Power System Faults, Detection and Classification

An electrical utility system is vulnerable to short circuit faults which can be caused by lines contacting each other or the ground due to lightning, vegetation, wind, earthquakes etc. [15]. These faults are categorized, with an approximate percentage of occurrences, mainly as single phase-to-ground faults (SPG, 70% - 80%), phase-to-phase-to-ground faults (DPG, 17% - 10%), phase-to-phase faults (PP, 10% - 8%) and three-phase faults (3PH, 3% - 2%) [15].

The fault detection and classification techniques make use of the changes in the current and voltage signals in case of a fault. The methods vary from hand-coded expert-defined rules based on certain thresholds to artificial intelligence-based techniques, such as ANNs, support vector machines, and fuzzy decision systems [2], [6]. Several features and transformations of the signals have been proposed and used for detection purposes like Fourier and wavelet transforms [6].

Although the protection of critical lines and system buses is ensured with local protection equipment like relays and circuit breakers, the data made available by PMUs offer the potential to increase the understanding and situational awareness in an energy management center as also proposed in [3] using the output of a PMU-only state estimator for fault detection and classification.

In this context, the approaches in [2] and [3] use decision trees, and [16] employs support vector machines for this purpose. Although the results presented in these works are promising, these approaches assume, as discussed above, complete presence of all the measurements in perfect synchronization.

In the scope of this work, we have experimented with two fault detectors for the output of a PMU-only state estimator: one based on ANN and the other based on support vector machines. Due to the observed superior performance of ANN and space limitations, we confine our discussion and results to fault detection and classification with ANNs in the following. Further work is ongoing for a comparison of different machine learning-based techniques for power system fault detection and classification.

A. *Artificial Neural Networks*

In pattern recognition and classification applications, ANNs, in particular the feed-forward back-propagation multilayer perceptrons (MLPs), have been widely used due to their outstanding performance [17]. As indicated in its name, an MLP consists of one or more hidden layers other than the input and the output layers. Each hidden layer has a non-linear activation or transfer function leaving the classifier capable of extracting the important features in the input data.

Based on the supervised back-propagation learning method, the MLP classifier utilizes a set of $N \in \mathbb{N}$ labeled input feature vectors, referred to as a training set, to adjust the weights of the neurons in its hidden layers. A training set $\mathbb{T}$ with N input-target pairs can be written as

$$\mathbb{T} = \{(\mathbf{x}_n, \mathbf{y}_n) \in \mathbb{R}^M \times \{0,1\}^K \mid n = 1, \ldots, N\}, \quad (1)$$

where $\mathbf{x}_n \in \mathbb{R}^M$ is the input feature vector, $\mathbf{y}_n = [y_{n1}, \ldots, y_{nK}]$ is the target output vector which refers to target class of the corresponding input vector $\mathbf{x}_n$, and K denotes the number of possible target classes. The target $\mathbf{y}_n$ is an all-zero vector except $y_{nk} = 1$ corresponding to the k^{th} class to which $\mathbf{x}_n$ belongs, where $k \in \{1, \ldots, K\}$.

In the current work, the input feature vector $\mathbf{x}_n \in \mathbb{R}^M$ is the output of the state estimator which consists of the per unit voltage magnitudes as well as the voltage angles of all the system buses. Note that, for a fault detection problem, K is equal to 2, referring to the occurrence of a fault or a normal operation. On the other hand, in case of fault classification, K is equal to the number of fault types vector $\mathbf{y}_n$ to represent possible faults, for example, SPG phase 1 (SPG1), DPG phases 1-2 (DPG12) etc.

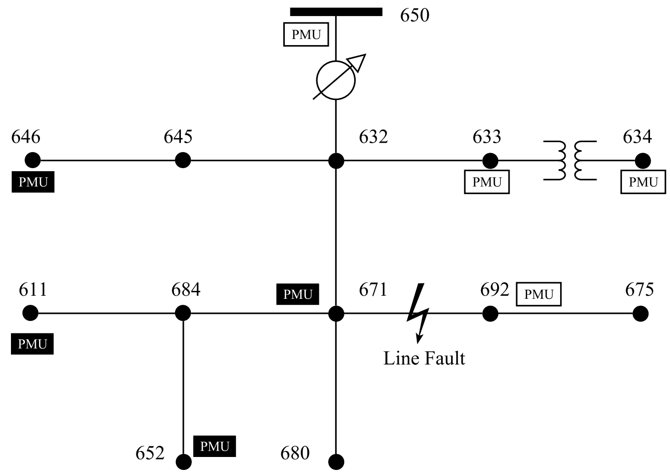

Fig. 3. The topology of the IEEE-13 bus test feeder [19]. A fault detector and classifier for the line between 671 and 692 is implemented and investigated. The PMUs shown in black and white boxes are assumed to send their data to different PDCs. A failure of black PMUs is also simulated.

In the training of ANN, the back-propagation approach updates the weights via the gradient descent method such that a certain cost function is minimized as further input vectors are introduced in the training. In this work, the utilized cost function during the training phase is the cross entropy, which is defined for the set $\boldsymbol{\Theta} = \{\mathbf{W}^{(\ell)} \mid \ell = 1, \ldots, L\}$ of weights for $L \in \mathbb{N}$ number of hidden layers as

$$J(\boldsymbol{\Theta}) = -\sum_{n=1}^{N} \sum_{k=1}^{K} y_{nk} \log \hat{y}_{nk}(\boldsymbol{\Theta}), \quad (2)$$

where $\mathbf{W}^{(\ell)} \in \mathbb{R}^{r_\ell \times r_{\ell-1}}$ is the weight matrix of the hidden layer ℓ, which has r_ℓ number of neurons, and $\hat{y}_{nk}(\boldsymbol{\Theta})$ is an estimate of y_{nk}, such that $\hat{y}_{nk}(\boldsymbol{\Theta}) = f(\boldsymbol{\Theta}, \mathbf{x}_n)$. In this content, $f(.)$ denotes the trained MLP model for either fault detection or classification problems.

Based on our experiments, a MLP classifier with three hidden layers each with 20 neurons has provided better performance given the training set. The *softmax* transfer function has been utilized at the output layer, whereas the *hyperbolic tangent sigmoid transfer function* is used as the activator in the hidden layers. For a comprehensive treatment of ANNs, please refer to [17], [18].

IV. Proposed Training Scheme and Simulations

This section presents the details of the proposed training scheme and its verification by simulation results. For the validation of the presented concept, we have carried out simulations of the IEEE 13-bus test feeder using the open source distribution system simulator OpenDSS [20] and generated the data for normal and faulty state conditions in a similar manner which is used in [21] and [22]. In the scope of this work, we aim to design a fault detector and a fault classifier for faults on the line between the buses 671 and 692. We use the neural network toolbox of MATLAB for this purpose. The topology of the test network is shown in Fig. 3. The fault resistances are simulated as $1\,\Omega$ and $5\,\Omega$. The normal and faulty state conditions are simulated under various load conditions, where the voltage and current phasor measurements are governed by zero-mean additive white Gaussian noise with an SNR of 30 dB. In order to generate the impact of the delayed measurements as illustrated in Fig. 2, measurement streams which consist of the shifted versions of the actual measurement values are generated. The measurements are then fed into a weighted least-squares estimator which delivers the estimates for the complex voltage phasors of all system buses [12]. The weighted least-squares state estimation is executed each milisecond with the available measurements at the SPDC.

Furthermore, in order to check the applicability of the proposed training approach, we generate two sets of data. The first set includes the output data of the state estimator when synchronized measurements for each time instance are used as the input feature and the corresponding system state as the target class, e.g. normal or faulty state in the training of the detector and the target fault type in the training of the fault classifier. In the second set, on the other hand, the measurements from different PMUs are assigned random communication delays drawn from a uniform distribution with the support $[d_{\min}, d_{\max}]$. In this way, training and test data sets are saved, which are generated by considering various load situations, various delay profiles, and fault resistances. The estimated system state at the output of the state estimator, resulting from the shifted measurements, are fed into the training as input feature vectors, where the system state is marked as faulty starting from the time instance at which the first measurement recorded just after the fault arrives at the SPDC in case of fault detection. Obviously, the detection of any fault is possible only after this moment. The trained neural network is then provided with test cases with various delay profiles in order to check their performance in terms of fault detection time and classification accuracy.

V. Results & Discussion

Fig. 4 and Fig. 5 illustrate the cumulative distribution function (CDF) of the fault *detection time* over all test cases (9000 test cases for each curve) for both considered training methods when the PMU communication delay interval $[d_{\min}, d_{\max}]$ is [10,80] and [10,180] ms, respectively. We define the *detection time* t_d as the duration between the occurence of the fault and the detection of the fault by the detector at the SPDC. The faulty operation is detected in all test cases successfully. The curves without markers show the results for the training with delayed data, whereas the curves with markers illustrate the training with the synchronized data. Furthermore, the dotted lines show the results for the cases when the PMU data originating from black PMUs are lost due to a PDC failure, see Fig. 3. Note that the stepwise CDF plots are due to the time step of state estimation, which is equal to 1 ms.

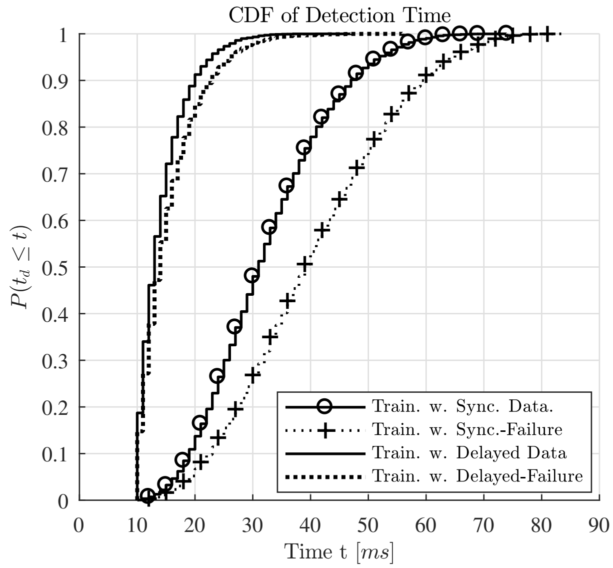

Fig. 4. CDF of the fault detection time in 9000 test cases when the communication delays are between 10 and 80 ms. Training the detector under consideration of time delays has the potential to improve the fault detection time compared with the training with synchronized data. Furthermore, in case of a PDC failure the degradation is relatively low in the proposed method.

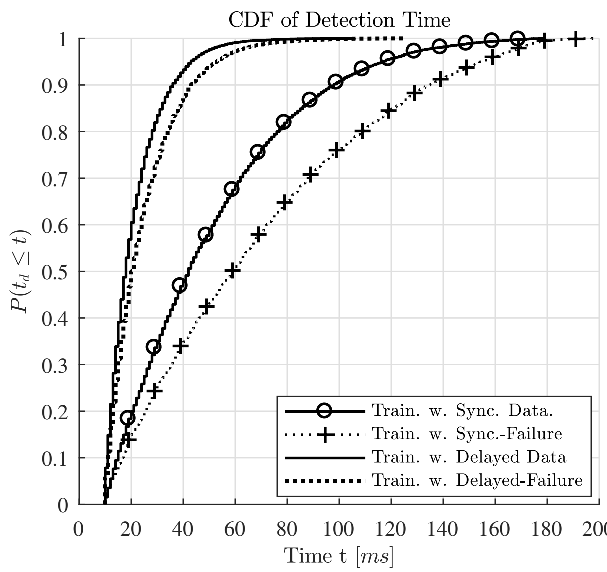

Fig. 5. CDF of the fault detection time in 9000 test cases when the communication delays are between 10 and 180 ms. The improvement is relatively more due to the higher variation in communication delays, see also Fig. 4

Referring to Fig. 4, we read that in 90% of all cases, the faulty condition is detected in less than 20 ms when the fault detector is trained under consideration of the delayed arrival of measurements, compared with only 20% of cases detected in less than 20 ms when trained with synchronized measurements. Similarly, we observe a considerable improvement in the tail of the distribution, i.e. in the maximum detection time.

Fig. 5 leads to similar conclusions with a relatively more significant improvement due to the larger variation of PMU delays in this case. When trained with delayed measurements, the detection time is less than 37 ms in 90% of all cases,

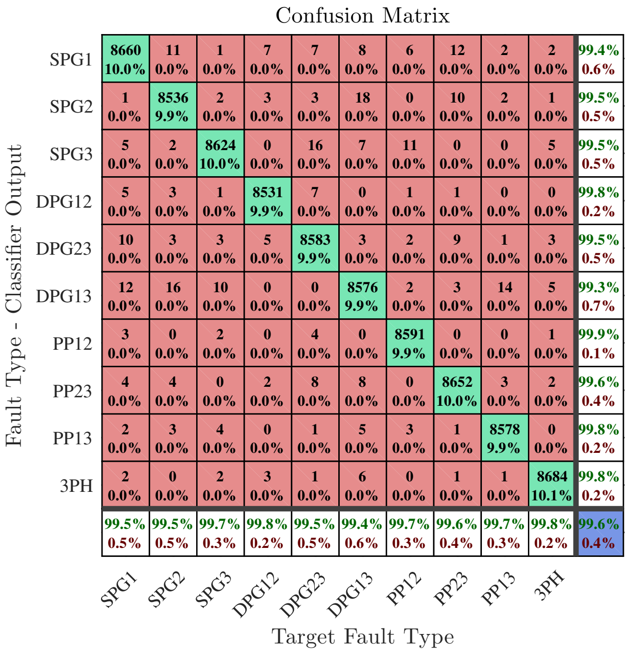

Fig. 6. Classification accuracy in percentage for all fault types when the communication delays are between 10 and 180 ms, see also Fig. 4. The last row in white background shows the correct classification rate for each type.

compared to a maximum detection time of 100 ms in 90% of all cases when trained with the synchronized measurements.

Moreover, in both figures, the degradation due to the loss of PMU data in failure cases, presented in dotted curves, is lower in the proposed training approach. Therefore, it can be said that the proposed approach is more robust in case of failures in the communication network.

Finally, Fig. 6 presents the classification accuracy of the implemented fault classifier in the form of a confusion matrix. The percentages of correct classification are shown along with the confusion rates for each fault type over all time instances which are detected as fault by the implemented fault detector. The white-colored last row in Fig. 6 reveals the rate of correct classification between 99.4% and 99.8% for all fault types.

VI. Conclusion

In this work, we have investigated the problem of fault detection and classification in power systems considering the communication network delays of phasor measurements. Unlike the available works in literature so far, which assume the presence of synchronized measurements in the detector and classifier, our proposed approach enables the detection and classification of faulty operations with the arrival of one or a part of impacted measurements. Therefore, the detection can be performed faster and reliably, also in case of failure or loss of some PMU measurements. The validity of the proposed approach has been verified through simulated data and the use of ANNs in detection and classification. Future work will consider the verification of the implemented method using real network or real-time simulation data. In addition, a comparison study of several machine learning techniques in terms of performance, complexity, and training time for power system fault detection and classification is under preparation.

References

[1] Final Report on the August 14, 2003 Blackout in the US and Canada: Causes and Recommendations. U.S.-Canada Power System Outage Task Force. Accessed:25.8.2017. [Online]. Available: https://energy.gov/sites/prod/files/oeprod/DocumentsandMedia/BlackoutFinal-Web.pdf

[2] D. Nguyen, R. Barella, S. A. Wallace, X. Zhao, and X. Liang, "Smart grid line event classification using supervised learning over pmu data streams," in *2015 Sixth International Green and Sustainable Computing Conference (IGSC)*, Dec 2015, pp. 1–8.

[3] F. Gao, J. S. Thorp, S. Gao, A. Pal, and K. A. Vance, "A voltage phasor based fault-classification method for phasor measurement unit only state estimator output," *Electric Power Components and Systems*, vol. 43, no. 1, pp. 22–31, 2015.

[4] O. Ardakanian, Y. Yuan, R. Dobbe, A. von Meier, S. H. Low, and C. Tomlin, "Event detection and localization in distribution grids with phasor measurement units," *CoRR*, vol. abs/1611.04653, 2016. [Online]. Available: http://arxiv.org/abs/1611.04653

[5] M. Jamil, S. K. Sharma, and R. Singh, "Fault detection and classification in electrical power transmission system using artificial neural network," *SpringerPlus*, vol. 4, no. 1, p. 334, Jul 2015.

[6] K. Chen, C. Huang, and J. He, "Fault detection, classification and location for transmission lines and distribution systems: a review on the methods," *High Voltage*, vol. 1, no. 1, pp. 25–33, 2016.

[7] K. S. Swarup and H. S. Chandrasekharaiah, "Fault detection and diagnosis of power systems using artificial neural networks," in *Proceedings of the First International Forum on Applications of Neural Networks to Power Systems*, Jul 1991, pp. 102–106.

[8] K. L. Butler and J. A. Momoh, "Detection and classification of line faults on power distribution systems using neural networks," in *Proceedings of 36th Midwest Symposium on Circuits and Systems*, Aug 1993, pp. 368–371 vol.1.

[9] "IEEE Standard for Synchrophasor Data Transfer for Power Systems," *IEEE Std C37.118.2-2011*, pp. 1–53, Dec 2011.

[10] H. A. Tokel, G. Alirezaei, and R. Mathar, "Integrated network design for measurement and communication infrastructures in smart grids," in *26th International Telecommunication Networks and Applications Conference (ITNAC 2016)*, Dunedin, New Zealand, Dec. 2016, pp. 267–273.

[11] "IEEE Guide for Phasor Data Concentrator Requirements for Power System Protection, Control, and Monitoring," *IEEE Std C37.244-2013*, pp. 1–65, May 2013.

[12] A. Abur and A. G. Exposito, *Power System State Estimation: Theory and Implementation*, ser. Power engineering. CRC Press, 2004.

[13] "EPRI Data Mining Initiative," http://smartgrid.epri.com/DMD-DMI.aspx, Electric Power Research Institute, (Accessed on 29.08.2017).

[14] "EPRI Data Analytics Case-Sequence of Outage Events Replay," http://smartgrid.epri.com/doc/DMD-Use-Cases/Sequence%20of%20Outage%20Events%20Replay%20-%20Short%20Version.pdf, Electric Power Research Institute, (Accessed on 29.08.2017).

[15] T. J. D. J. Lewis Blackburn, *Protective Relaying: Principles and Applications*. CRC Press, 2014.

[16] D. Nguyen, S. Wallace, and X. Zhao, "Finding needles in a haystack : Line event detection on smart grid pmu data streams," in *2016 Sixth International Conference on Smart Grids, Green Communications adnd IT Energy-aware Technologies*, 2016, pp. 42–47.

[17] S. Haykin, *Neural Networks. A Comprehensive Foundation*, 2nd ed. Pearson Education, 1999.

[18] K. Murphy, *Machine Learning: A Probabilistic Perspective*, 2nd ed. MIT Press, 2012.

[19] "Distribution Test Feeders - Distribution Test Feeder Working Group," http://www.ewh.ieee.org/soc/pes/dsacom/testfeeders/index.html, IEEE Power & Energy Society, (Accessed on 04.09.2017).

[20] R. C. Dugan, "Reference guide: The open distribution system simulator (OpenDSS)," *Electric Power Research Institute, Inc*, vol. 7, 2012.

[21] J. Cordova and M. O. Faruque, "Fault location identification in smart distribution networks with distributed generation," in *2015 North American Power Symposium (NAPS)*, Oct 2015, pp. 1–7.

[22] A. Anwar, A. N. Mahmood, and Z. Shah, "A data-driven approach to distinguish cyber-attacks from physical faults in a smart grid," in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 2015, pp. 1811–1814.