Feature-Selection based PMU Placement for Detection of Faults in Power Grids

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

***Abstract*—The monitoring of power distribution networks and identification of system faults become more and more important as the overall power grid undergoes structural changes. Such changes are due to the increasing integration of distributed and volatile renewable generation units. This work focuses on strategies for the placement of phasor measurement units (PMU) in a power distribution system, such that the detection and classification of line outages can be facilitated accurately.**

The determination of sensor locations is based on a feature selection approach, where the measurement locations, which provide the most informative input to the supervised learning techniques, are successively selected until the required classification accuracy is obtained. Hence, the number of required PMUs is minimized without jeopardizing the detection accuracy. The proposed methodology is applied to benchmark distribution systems, where simulated data is used for training deep neural networks (DNN), decision trees (DT), and random forests (RF) for fault detection and classification.

I. Introduction

Situational awareness is a key requirement for power system operators to provide a reliable service to their customers. As the overall power grid transforms from a centralized system to a distributed one along with the integration of even increasing number of small-size and volatile renewable generation units, particularly distribution networks are expected to play a crucial role in the overall ecosystem of a smart grid. As a result, distribution system operators have to comply with more and stricter regulations for a better controlled grid with less system interruptions, which are evaluated in terms of service quality indicators, such as the system average interruption duration index (SAIDI) [1].

In this context, the detection of system faults and the identification of their locations in the network are important practical problems. The best usage of both the available sensor information and equipment status is crucial for an automated detection and identification, which are required for further actions in the reconfiguration of the whole network. Nevertheless, the distribution systems have not in general been equipped with extensive monitoring equipment due to their passive role in the traditional power grid architecture. For this reason, many distribution operators put efforts into the gradual roll-out of measurement units and the required communication links in their systems [2], [3]. Furthermore, the number of sensors in the distribution networks is expected to increase also as a result of the ongoing efforts to develop low-cost micro PMUs (μPMU) for distribution level applications, see for example [4].

The proper selection of sensor locations for fault detection has been widely studied for various engineering systems, where two main methods can be identified in the literature. The first group of methods relies on a mathematical or a structural model of the system under consideration, see for example [5]. The drawback of these methods is the requirement of an accurate system model which can also describe the system behavior under faulty conditions. The second group of methods utilizes data-driven approaches to achieve an acceptable detection performance. For example, [6] proposes a sensor placement technique based on fuzzy feature selection for the detection of faults in the process of pharmaceutical synthesis in a heat-exchanger reactor. There are two main advantages of the data-driven approaches. Firstly, they are model independent, and secondly, they inhere the potential to identify further pattern changes caused by system faults, which would otherwise have not been captured by the system model. Therefore, the data-driven methods can be used to design strategies for the selection of the minimum required number of sensors by means of techniques such as feature selection and dimensionality reduction.

In the domain of power systems, the latter approach has been applied to the problem of sensor placement with the objective of minimization of state estimation errors targeting at voltage values of key system buses, see for example [7]. On the other hand, the works in the literature, which deal with the placement of sensors for the detection of faults, have taken approaches based on the grid topology. For instance, [8] and [9] propose integer linear programming models for the selection of a minimum number of PMU locations. For example, the optimization problem in [9] is constructed based on the assumption that there must be at least one PMU at either end of any line.

The contribution of the present work is a data-driven strategy for the placement of PMUs using feature selection methods, to achieve the desired accuracy of fault detection within the network. In addition, based on the optimal placement of PMUs, the performance of machine learning based detection methods have been evaluated.

The rest of this paper is organized as follows: Section II presents the feature selection methods investigated in this paper to choose the most relevant measurements as an input

to the detector. Section III introduces the machine learning-based fault detection methods utilized in this work. The details of the simulation study, as well as the simulation results, are presented in Section IV.

II. Feature Selection Techniques

The first step in any classification problem is to select a subset of the available features, which contains the most influential and characterizing ones. In the current work, the available feature set consists of the measurement values which can potentially be provided by the PMUs. Such PMUs can be installed at the selected system nodes. In general, a PMU, which is installed at a certain system node, can provide a subset of the available features, namely *i)* bus voltage at its node, *ii)* currents and power flows on all incident lines.

Note that a proper reduction in the dimensionality of the feature space can eliminate the redundancy in the data, reduce the complexity of the required classifier, and improve the learning process which would lead to an improvement of the classification performance and a reduction in the training time. Furthermore, it can also be used to determine the required number and the optimal locations of the PMUs for an accurate detection of the fault location, as shown in the present work.

For an optimal selection of features, three methods are considered in this work, namely variance thresholding (VT), correlation coefficient (CC) measure, and mutual information (MI) test.

The concept behind VT is that the features with low variance contain basically less information, and contribute less to a successful classification. Therefore, VT eliminates features with a variance lower than a fixed threshold, and thus preserves the most valuable features in the feature space, which leads to a more stable performance [10], [11]. In the current work, rather than defining an optimal threshold, the features are simply ranked by their variances, since the selection of features will follow iteratively starting from the best feature until a satisfactory classification accuracy is guaranteed.

The CC measure of a certain feature vector is defined as the covariance between the vector and the corresponding class vector divided by the multiplication of their individual standard deviations. The CC measure takes a value not bigger than 1, which indicates highly relevant and important features, whereas 0 indicates irrelevant features which are to be neglected.

Finally, the MI measure from information theory can be utilized to perform feature selection. The MI is a measure of the mutual dependency between two random variables [12]. Let A and B be two random variables. Their MI measure, denoted by $I(A;B)$, is written as

$$I(A;B) = \sum_{a\in A}\sum_{b\in B} p(a,b)\log\frac{p(a,b)}{p_A(a)\ p_B(b)}, \qquad (1)$$

where p_A and p_B denote the probability density function (PDF) of A and B, respectively, and $p(a,b)$ denotes their joint PDF. If A and B are independent, then $I(A;B)=0$. In the case that A and B belong to unknown PDFs and they have N_A and N_B samples, respectively, one can estimate their PDFs in (1) by sorting the measurements (points) into k clusters based on the k-nearest neighbor distance as explained in [13]. If $z_A(i)$ and $z_B(j)$ denote the number of points in cluster i and j, and $z_{A,B}(i,j)$ denotes the number of points falling in their intersection, then, $p_A(i) \approx \frac{z_A(i)}{N_A}$, $p_B(j) \approx \frac{z_B(j)}{N_B}$, and $p(i,j) \approx \frac{z_{A,B}(i,j)}{N_A+N_B}$ [13].

For the problem in hand, MI provides a measure of the information shared between each feature and the class label [14]. As a result, the set of features are ranked based on their MI values. Whenever this measure equals zero, it can be concluded that this certain feature and the target vectors are independent, hence this feature is irrelevant for classification purposes and can be excluded [15].

In order to choose the most appropriate feature selection method, and to provide the detection accuracy given the measured quantities being ranked via the previously mentioned feature selection methods, two classification methods have been utilized, namely the decision tree (DT) and deep neural networks (DNN). For fault detection purposes, random forest (RF) is tested in addition to DT and DNN. The three detection methods are explained in the following section.

III. Machine Learning-Based Fault Detection

Machine learning approaches and data-driven techniques provide the possibility to automate the power grid and its processes. Among different applications, identifying the location of faults in power grid, e.g., line outages, is an issue which can be directly addressed and solved using data-driven methods as a multi-class classification problem. In the present work, DTs, RFs, and DNNs are chosen as the classifier models.

A. Deep Neural Network (DNN)

Due to the rapid development in the field of artificial neural networks (ANNs), more attention is drawn to deep neural networks (DNNs) for their outstanding performance in classification, regression, and pattern recognition tasks. The feedforward DNNs are implemented with multiple hidden layers, each of which contains a specified number of neurons with certain weights, loosely similar to the neurons in the human brain. The weighted input is summed over all layers and transformed through the activation functions of individual neurons to the output layer [16]. As a supervised machine learning approach, a DNN is trained such that the cross-entropy loss function $\mathcal{L}(f(\mathbf{X},\mathbf{\Theta}),\mathbf{t})$ is minimized, where $\mathbf{X}=[\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_N]^T \in \mathbb{R}^{N\times M}$ is the training set of $N \in \mathbb{R}$ samples, with $\mathbf{x}_i \in \mathbb{R}^{M\times 1}$ is the measurement vector containing $M \in \mathbb{R}$ features of sample $i \in \{1,\ldots,N\}$. Furthermore, f denotes the function for the output of the DNN, i.e., the estimated classes, and $\mathbf{t} \in \mathbb{R}^{N\times 1}$ is the actual classes of the corresponding N samples. The set $\mathbf{\Theta}=\{\mathbf{W}^{(d)} \mid d=1,\ldots,D\}$ represents the weights of the $D \in \mathbb{N}$ hidden layers, where $\mathbf{W}^{(d)} \in \mathbb{R}^{r_d\times r_{d-1}}$ is the weight matrix of the d^{th} layer, which has r_d neurons [17]. The implemented DNN has been designed with four fully connected layers, each with 50 neurons. The activation function of the

output layer has been implemented as the softmax function, which is widely used for a multi-class classification problem [16].

B. Decision Tree (DT)

A decision tree (DT) is a classification algorithm, which is best used to classify data with common attributes [18]. Just as the name indicates, this supervised learning classifier has a similar structure to a tree, such that it has a set of nodes, edges and leaves. An edge is a connection of at least two nodes, or a node and a leaf. Whenever the DT decides to separate or classify the data based on a certain feature, a node is generated. The number of edges in a node indicates the possible values, that a certain feature can have. On the other hand, the leaves are the outputs of the classification and indicate the estimated output class of the input sample [19].

C. Random Forest (RF)

Random forest is an example of an ensemble approach, as in specific bootstrap sampling. The underlying idea is to combine a group of DTs in order to construct a more powerful and more robust learner against over-fitting [20]. The main characteristic of a RF is that it is built based on randomness. Unlike DT, the split in RF is based on the best among a subset of randomly chosen predictions at that node [21], [22]. The main idea behind the randomness is to decorrelate the trees, such that the resulting ensemble scheme has a low variance [23].

IV. Application on Test Networks

A. Power Grid Data Aggregation

Acquiring clean and accurate data is an important step and has great influence on the resulting performance. Different benchmark grid models have been considered in this work. Upon defining the grid under consideration, GridLAB-D simulator is utilized for power flow simulations and for extracting the measurements [24].

For each of the considered grid scenario, one power source is available. From this source, we have a grid with a radial structure, such that whenever a fault occurs, a sub-scenario is generated. Therefore, the total number of sub-scenarios N_F equals ℓ, the number of line segments, which are investigated for the detection of faults. Hence, there are $K = N_{\mathrm{F}} + 1$ target classes in the classification process with the addition of the state of normal operation to the N_F faulty states.

In the following, we briefly introduce the steps of data generation. After defining the normal operation, i.e., no fault is occurred, each of the $\ell \in \mathbb{N}$ lines is alternatingly considered as the faulty line one at a time. The simultaneous occurrence of multiple faults is not considered at this point. Furthermore, each scenario is independently simulated without any transient behaviour.

Afterwards, each sub-scenario is simulated separately to obtain a set of measurements, where the values of the connected loads are varied between 0 and 150% of their default values in a uniform and random manner. The recorded measurements are obtained by placing PMUs at all nodes to measure the following features: the real and imaginary parts of voltage phasors at all n nodes, *and* the active and reactive part of the power flows in all ℓ lines. Thus, the total number of features collected from the whole grid is equal to $M = 2(n + \ell)$. Per sub-scenario, i.e., class, $S \in \mathbb{N}$ samples are generated, such that $N = S \cdot K$, and $N \gg N_{\mathrm{F}}$.

Two well-known and widely utilized benchmark grids have been considered in this work, namely IEEE 13-node test feeder and IEEE 37-node test feeder [25]. Some minor modifications have been made to the two grids without changing their structure. For example, the IEEE 13 feeder contains 12 links to connect 13 nodes. However, for simulation purposes, two extra nodes are added to avoid any errors in the power flow simulator GridLAB-D. One node is added after the voltage regulator at node 630. The other one is added between the nodes 632 and 671 to introduce an extra line. Hence, in total, the simulated 13-node feeder contains $n = 15$ nodes and 14 links. Furthermore, among the 14 links, the following ones are not considered as faulty in the simulation:

- the transformer between nodes 633 and 634,
- the switch between nodes 671 and 692, and
- the voltage regulator between nodes 650 and 632.

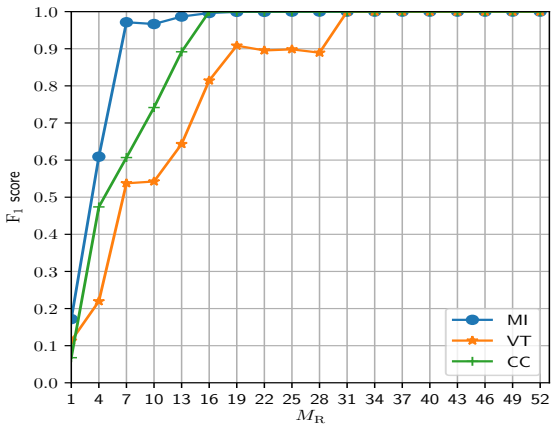
Thus, the simulated 13-node feeder contains $n = 15$ nodes and $\ell = 11$ lines. This leads to $K = 12$ classes or sub-scenarios representing 11 fault locations and the normal operation case, and $M = 52$ features can be measured or generated from the grid.

Similarly, one node has been added to the IEEE 37-node feeder after the voltage regulator at node 701. Both the available transformer *and* the voltage regulator are not considered as faulty among the cases. Thus, the 37-node feeder has been considered with $\ell = 35$ lines, and $n = 38$ nodes, which means $K = 36$ classes or sub-scenarios are to be detected including the normal operation. The total number M of measured quantities is 146. In both grid models, $S = 14000$ examples are generated from the simulation environment per class or sub-scenario.

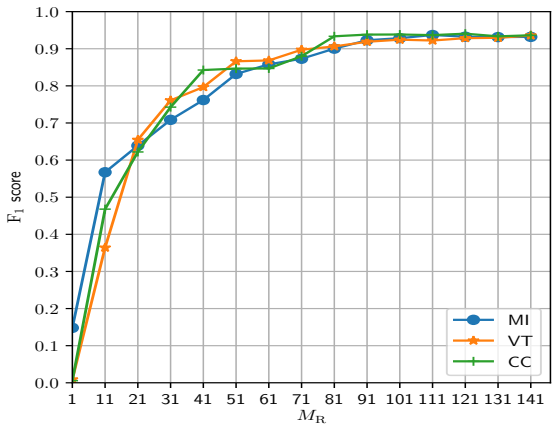
B. PMU Placement Based on Feature Selection Results

The PMU locations are sequentially selected according to the sequence of the selected features which are ranked based on the feature selection procedure. In order to ensure an efficient utilization of all placed PMUs, all the quantities, which are measured or recorded by the placed PMU, are also included based on their ranking according to the used metric. By this approach, the number of the required PMUs can further be decreased.

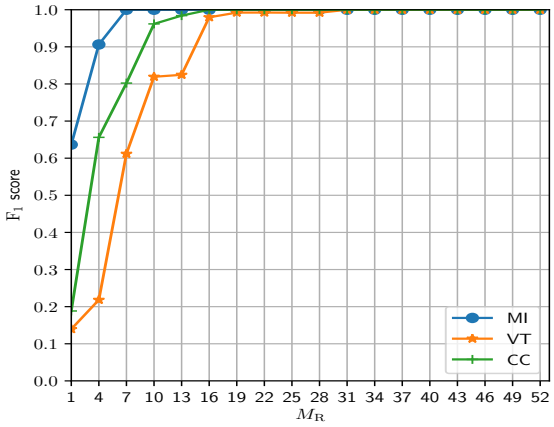
The F_1 score [26] has been considered as a suitable metric for the classification accuracy. Figure 1 presents the F_1 score for the two grids as a function of the number of features based on the feature selection methods and the DNN and DT detection methods. It can be observed that the optimal feature selection method is a function of the measured quantities from the grid itself. On one hand, in the 13-node feeder, a feature selection approach via the MI method reveals the best $M_{\mathrm{R}} = 7$ relevant features, i.e., with 13.5% grid coverage, to



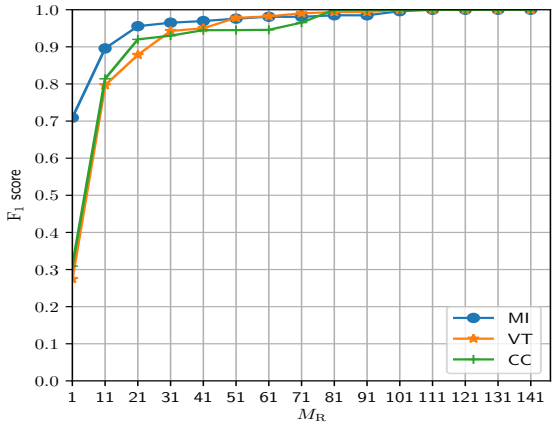
(a) IEEE 13-Node Test Feeder with fault detection by DNN.



(b) IEEE 37-Node Test Feeder with fault detection by DNN.



(c) IEEE 13-Node Test Feeder with fault detection by DT.



(d) IEEE 37-Node Test Feeder with fault detection by DT.

Fig. 1. F_1 score evaluated by the iterative selection of features ranked by VT, CC, and MI measures, for the two considered benchmark models depending on the number of features included in the classification process.

achieve an F_1 score close to 1 via the DT detection method, as shown in Figure 1(c). Hence, in the 13-node feeder, 2 PMUs, which provide the most relevant $M_{\mathrm{R}} = 7$ features, suffice to achieve an accurate detection. Similar results are obtained using the MI test via the DNN detection method, as shown in Figure 1(a). However, the VT approach requires, for a similar detection performance, a higher grid coverage with the minimum required number of relevant features as 20 and 32 for DT and DNN, respectively.

On the other hand, the considered feature selection methods exhibit asymptotically similar performances for each individual detection method for the 37-node feeder. For example, in Figure 1(d), F_1 score of 0.958 has been achieved via DT using the three feature selection methods with $M_{\mathrm{R}} = 48$ features. This corresponds to a grid coverage of 33% and achieved using 10 PMUs. A relatively lower performance is observed in Figure 1(b) via the DNN classification method, where an F_1 score of 0.9 is attained using the three feature selection methods. To achieve this performance, $M_{\mathrm{R}} = 80$ features are needed, i.e., a grid coverage of 55% is required. Hence, the 37-node feeder requires 20 PMUs according to the DNN classification method.

The MI score has been chosen to be the appropriate feature selection method to be utilized for both grid models, as it shows a relatively better performance than the CC and VT methods. In other words, the detection process utilizes the ranked features according to the MI score.

Figure 2 presents the most relevant quantities in the 37-node feeder, which correspond to 10 required PMUs to achieve an F_1 score of 0.958 via the DT method. Note that, as previously discussed, the node 781 has been added to the reference grid for simulation purposes. For both test feeders, the process of adding features to improve the performance is terminated when F_1 score is greater than 0.95 and the relative improvement of the F_1 score is less than 0.1% with the addition of a new feature. This process reveals, at the end, the required number

TABLE I
CLASSIFICATION ACCURACY FOR THE IEEE 13-NODE FEEDER

Classification method	Classification Accuracy for Training Data Set	Test Data Set
DT	1.0	1.0
RF	1.0	1.0
DNN	0.988	0.987

TABLE II
CLASSIFICATION ACCURACY FOR THE IEEE 37-NODE FEEDER

Classification method	Classification Accuracy for Training Data Set	Test Data Set
DT	0.958	0.958
RF	0.958	0.958
DNN	0.963	0.955

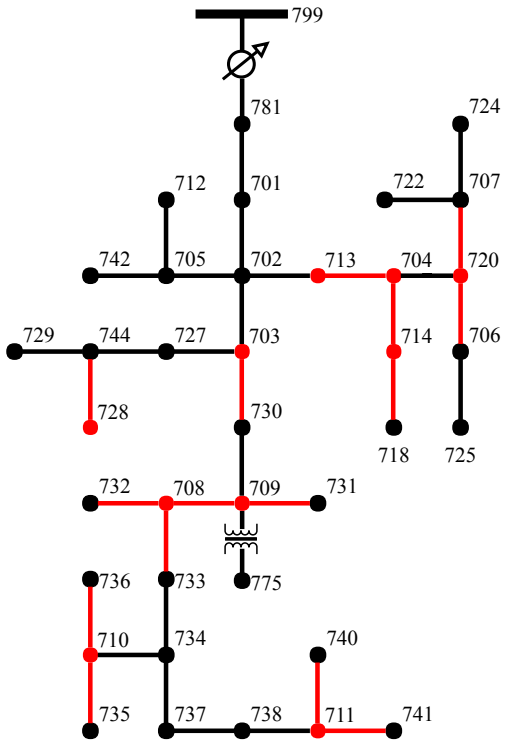


Fig. 2. The most relevant quantities to be measured to achieve a detection accuracy of 0.958 in the IEEE 37 grid model. The red nodes and lines indicate the required voltage measurement at the nodes and the required power flow measurements on lines, respectively.

of features.

C. Fault Detection Results

Upon deciding the optimal placement of the PMUs as the output of the feature selection stage, the detection process takes place by training the classifiers on the subset of most relevant features M_{R}. In other words, the considered data set for detection, i.e., classification is $\mathbf{X} \in \mathbb{R}^{N \times M_{\mathrm{R}}}$, instead of a training matrix of size $N \times M$.

Table I and Table II present the F_1 scores for the previously presented classification methods, namely DT, RF and DNN. Simulation results have shown that the performances of the RF and the DT classifiers are very close to each other, and outperform the DNN. One explanation can be that the utilized grids have a radial structure which is similar to the structure of those classifiers. Furthermore, the process of choosing the optimal structure of the DNN in terms of number of layers and neurons in each layer is unfortunately not a straightforward task. On the other hand, the fitting of non-parametric RF and DT classifiers is less challenging, which can be a reason in their superior performance.

V. CONCLUSION

In this work, we have proposed a strategy based on data analysis and feature selection for the placement of a minimum number of required PMUs to facilitate an accurate detection of line outages in power grids. The optimal selection of the locations of the PMUs is based on the identification of the relationship of available measurements with the target class vector by the metrics of variance, mutual information, and correlation coefficient. The proposed approach enables not only the determination of the required optimal number of PMUs, but it also decreases the complexity of the required classifier and the training time as a result of eliminating redundant and less relevant features. The resulting performances of the classifiers, namely deep neural networks, decision trees, and random forests, have been evaluated through extensive simulations, where decision trees and random forest classifiers have outperformed the deep neural network classifier. The results presented in this paper show that certain data patterns in different cases of normal or faulty operation can be recognized and exploited in order to decrease the number of PMUs while still ensuring the required accuracy of detection.

ACKNOWLEDGMENT

This work is supported by the European project *Service Oriented Grid for the Network of the Future (SOGNO)* under the umbrella of *Horizon 2020* with the grant agreement number 774613. Furthermore, the authors would like to thank Dr. Arash Behboodi for valuable inputs and discussions.

REFERENCES

[1] "Trends in electricity distribution network regulation in North West Europe," http://energinorge.nsp01cp.nhosp.no/getfile.php/FILER/NYHETER/NETT OG SYSTEM/trends in electricity distributon network regulation in north west europe.pdf, Frontier Economics Ltd, London, Mar 2012, (Accessed on 01.08.2018).

[2] J. Chai, Y. Liu, J. Guo, L. Wu, D. Zhou, W. Yao, Y. Liu, T. King, J. R. Gracia, and M. Patel, "Wide-area measurement data analytics using FNET/GridEye: A review," in *2016 Power Systems Computation Conference (PSCC)*, June 2016, pp. 1–6.

[3] H. A. Tokel, G. Alirezaei, and R. Mathar, "Integrated network design for measurement and communication infrastructures in smart grids," in *2016 26th International Telecommunication Networks and Applications Conference (ITNAC)*, Dec 2016, pp. 258–264.

[4] A. Angioni, G. Lipari, M. Pau, F. Ponci, and A. Monti, "A low cost PMU to monitor distribution grids," in *2017 IEEE International Workshop on Applied Measurements for Power Systems (AMPS)*, Sept 2017, pp. 1–6.

[5] M. Krysander and E. Frisk, "Sensor placement for fault diagnosis," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 38, no. 6, pp. 1398–1410, Nov 2008.

[6] L. Hedjazi, T. Kempowsky-Hamon, L. Despnes, M. L. Lann, S. Elgue, and J. Aguilar-Martin, "Sensor placement and fault detection using an efficient fuzzy feature selection approach," in *49th IEEE Conference on Decision and Control (CDC)*, Dec 2010, pp. 6827–6832.

[7] M. Ferdowsi, B. Zargar, F. Ponci, and A. Monti, "Design considerations for artificial neural network-based estimators in monitoring of distribution systems," in *2014 IEEE International Workshop on Applied Measurements for Power Systems Proceedings (AMPS)*, Sept 2014, pp. 1–6.

[8] S. P. Pokharel and S. Brahma, "Optimal PMU placement for fault location in a power system," in *41st North American Power Symposium*, Oct 2009, pp. 1–5.

[9] S. S. Geramian, H. A. Abyane, and K. Mazlumi, "Determination of optimal PMU placement for fault location using genetic algorithm," in *2008 13th International Conference on Harmonics and Quality of Power*, Sept 2008, pp. 1–5.

[10] Y. Han and L. Yu, "A variance reduction framework for stable feature selection," in *2010 IEEE International Conference on Data Mining*, Dec 2010, pp. 206–215.

[11] J. Van Hulse, T. M. Khoshgoftaar, A. Napolitano, and R. Wald, "Threshold-based feature selection techniques for high-dimensional bioinformatics data," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 1, no. 1, pp. 47–61, Jun 2012. [Online]. Available: https://doi.org/10.1007/s13721-012-0006-6

[12] S. Khalid, T. Khalil, and S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning," in *2014 Science and Information Conference*, Aug 2014, pp. 372–378.

[13] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Phys. Rev. E*, vol. 69, p. 066138, Jun 2004. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.69.066138

[14] C. Krier, F. Cois, V. Wertz, and M. Verleysen, "Feature Scoring by Mutual Information for Classification of Mass Spectra," in *FLINS 2006, 7th International FLINS Conference on Applied Artificial Intelligence*, Genova (Italy), 2006, pp. 557–564.

[15] J. R. Vergara and P. A. Estévez, "A review of feature selection methods based on mutual information," *Neural Computing and Applications*, vol. 24, no. 1, pp. 175–186, Jan 2014. [Online]. Available: https://doi.org/10.1007/s00521-013-1368-0

[16] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. The MIT Press, 2016.

[17] H. A. Tokel, R. Alhalaseh, G. Alirezaei, and R. Mathar, "A new approach for machine learning-based fault detection and classification in power systems," in *2018 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, Washington DC, USA, 2018.

[18] Z.-S. Pan, S.-C. Chen, G.-B. Hu, and D.-Q. Zhang, "Hybrid neural network and C4.5 for misuse detection," in *Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.03EX693)*, vol. 4, Nov 2003, pp. 2463–2467 Vol.4.

[19] J. Leskovec, A. Rajaraman, and J. D. Ullman, *Mining of Massive Datasets*, 2nd ed. New York, NY, USA: Cambridge University Press, 2014.

[20] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct 2001. [Online]. Available: https://doi.org/10.1023/A:1010933404324

[21] S. S. Baboo and E. Iyyapparaj, "A classification and analysis of pulmonary nodules in CT images using random forest," in *2018 2nd International Conference on Inventive Systems and Control (ICISC)*, Jan 2018, pp. 1226–1232.

[22] A. Liaw and M. Wiener, "Classification and regression by random forest," *R News*, vol. 2, no. 3, pp. 18–22, 2002. [Online]. Available: https://CRAN.R-project.org/doc/Rnews/

[23] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: Data mining, inference, and prediction*. Springer Series in Statistics, 2001.

[24] D. P. Chassin, K. Schneider, and C. Gerkensmeyer, "GridLAB-D: An open-source power systems modeling and simulation environment," in *2008 IEEE/PES Transmission and Distribution Conference and Exposition*, April 2008, pp. 1–5.

[25] IEEE Power and Energy Society. (2018) IEEE PES Test Feeders. [Online]. Available: sites.ieee.org/pes-testfeeders/resources/

[26] K. Saleh, M. Hossny, and S. Nahavandi, "Driving behavior classification based on sensor data fusion using LSTM recurrent neural networks," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, Oct 2017, pp. 1–6.