Transmission Line Outage Detection with Limited Information Using Machine Learning

Daniel Flores, Yuanrui Sang, and Michael P. McGarry
Department of Electrical and Computer Engineering
The University of Texas at El Paso
El Paso, TX, USA
dflores45@miners.utep.edu, ysang@utep.edu, mpmcgarry@utep.edu

2023 North American Power Symposium (NAPS) | 979-8-3503-1509-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/NAPS58826.2023.10318637

***Abstract*—Transmission line outage detection plays an important role in maintaining the reliability of electric power systems. Most existing methods rely on optimization models to estimate the outage of transmission lines, and the process is computationally burdensome. In this study, we propose a transmission line outage detection method using machine learning. Using this method, we could monitor the power flow of one line and estimate whether another line is in service or not, despite the load fluctuations in the system. The study also investigates the principles for observation point selection and the effectiveness of this method in detecting the outage of transmission lines with different levels of power flows. The method was implemented on an IEEE 118-bus system, and results show that the method is effective for transmission lines with all levels of power flows, and line outage distribution factors (LODF) are good indicators in observation point selection.**

***Index Terms*—KNN, line outage distribution factors (LODF), machine learning, transmission faults, transmission outage detection**

I. Introduction

The electric power grid is one of the critical infrastructures supporting our daily life, and its reliability is paramount. According to a report by Lawrence Berkeley National Laboratory, power outages cause an economic loss of $79 billion annually in the U.S. [1], and reducing power outages can result in significant improvements in social welfare. The transmission network is a critical power grid component, since it delivers electric power over long distances, and transmission outages can have a series of negative consequences in power system operations, such as cascading failures and power outages [2], [3]. Thus, detecting transmission line outages can have a tremendous positive impact on enhancing power system reliability.

Various reasons, such as lightning, tree fall, or extreme weather, can cause transmission outages. Among these reasons, extreme weather can cause the most widespread transmission failure. For example, during Hurricane Ian, the Eastern Interconnection reported 140 transmission outages [4]. With the wide installation of phasor measurement units (PMU), an increasing amount of data is available for power system monitoring. However, transmission outage detection is a computationally complex problem [2], [5], [6], and the computational efficiency of such problems can be improved through a distributed approach [7], [8], graph theory [9], or innovative circuit theory-based method [10]. Transmission outage detection is challenging even when complete information about power system operation is available, and the unavailability of system data under certain circumstances adds to the difficulty of this problem. Because of the high cost of PMU, not all the substations are installed with PMUs [11]. Also, during natural disasters, not only the power grid gets damaged but also the communication network. This requires us to identify the transmission outages without complete information of the system. Existing transmission outage detection method are mostly optimization-based [12], [13], [14], and the locations of PMUs can be selected considering the effectiveness of transmission outage detection [15]. Although PMUs do not have to be installed at every bus, at least almost half of the buses in the system need to be installed with PMUs to ensure accurate monitoring in such methods [16]. An emerging method to improve the computational efficiency and effectiveness of transmission outage detection is through machine learning. A data-driven topology identification method is proposed in [17], and using this method, to ensure high detection accuracy, measurements at a large number of nodes in the system is desired. The topology of a radial distribution network is estimated using machine learning in [18], which shows the potential of machine learning in network topology estimation.

Currently, there is still a gap in detecting transmission outages with high computational efficiency with limited information. This study aims to address this gap by proposing a machine learning-based transmission outage detection, which can detect transmission outages with high speed and limited information. Using this method, we can monitor the power flow on one transmission line considering the uncertainty of load and estimate whether another transmission line is in or out of service. The method was implemented on an IEEE 118-bus system, with power flow data collected through the ACOPF solver of MATPOWER [22] and the learning process performed by the `Scikit-learn` machine learning library. The case study investigates the principles for observation point selection, i.e., which transmission line is the best to be monitored to estimate the outage of the other, and the results show that line outage distribution factor (LODF) is a good indicator for observation point selection.

This paper is organized as follows. Section II describes the

This work was supported by the U.S. National Science Foundation award #2131201.

proposed method, Section III our experiment design, Section IV our experiment results and findings. Finally, Section V concludes the paper and outlines avenues for further investigation.

II. Detecting Transmission Line Outages

Machine learning (ML) can be applied to detect power system transmission line outages. A binary nominal variable can be used to represent the status for each transmission line: 0 - line is *in-service* and 1 - line is *out-of-service*: y_i where i is the transmission line number. Collectively those variables form a vector $\mathbf{y}$ of length M where M is the number of transmission lines in the power system. If $M = 100$, there are 100 transmission lines and we use ML to predict the values of 100 binary nominal variables. With this modeling approach we can detect any number of simultaneous line outages. However, we can model the problem in other ways that do not have that ability. As an example, if we needed to only detect a single line outage we could model the problem using a single nominal variable, y, that would assume an integer value of $0 \ldots M$; e.g., 0 indicating no line outage, 1 indicating an outage on line 1, and M indicating an outage on line M.

Predicting the value of a nominal variable is referred to as *classification* and there are many ML algorithms for classification, such as logistic regression, k-nearest neighbors (KNN), and decision trees [21]. The input to our specific classification problem would be data collected from a set of *observation points* in the power system. The collected data could be real and reactive power measurements, voltages, and currents observed at each member of the set of observation points. These data would be referred to as *features* in ML terminology; these power system features are used by the classification model to predict the binary nominal variables indicating the line outages. Let $x_{i,j}$ be feature j collected at observation point i and $\mathbf{X}$ be the matrix of those values, all features from all observation points. The classification model is a function that maps the features to the binary nominal variables representing the state of the transmission lines,

$$\mathbf{y} = \mathcal{F}(\mathbf{X}) \tag{1}$$

An ML classification algorithm, e.g., KNN, produces a function $\mathcal{F}$ that minimizes error between that function and training data (i.e., matching pairs of $\mathbf{y}$ and $\mathbf{X}$) provided to the algorithm. This is a slight simplification, see [21] for more detail on classification with ML.

In summary, the line outage detection system contains the following steps, see Fig. 1:

1) collect data from the observation point(s)
2) use ML classification model to infer the binary nominal variables representing the line outage state of the system
3) send indications regarding the location(s) of line outage(s)

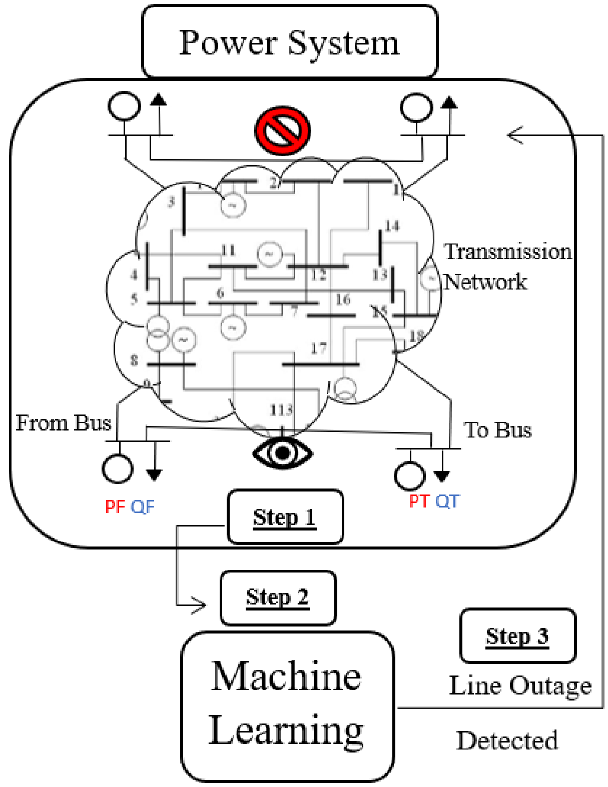

Fig. 1. Flow Chart of Transmission Line Outage Detection

In this work, we gain some insight into the selection of power system observation points that provide the features to be used for the line outage detection classification problem. Specifically, we gain insight into the observation points best for observing real power at the *from* buses (PF or $x_{i,1}$), reactive power at the *from* buses (QF or $x_{i,2}$), real power at the *to* buses (PT or $x_{i,3}$), and reactive power at the *to* buses (QT or $x_{i,4}$). With four features to be collected at each observation point we want to determine the best selection of observation points to collect those features. The goal is to provide sufficient classification performance as measured by precision, recall and F-1 score. Here we consider the use of a single observation point and leave the consideration of two or more observation points for future work. In this study, we choose the observation points based on the LODF values [19], [20] of observed lines versus the line out of service. A high LODF indicates that the observed line will experience a relatively large change in its power flow when the other line goes out of service, and thus, we hypothesize that lines with high LODF values corresponding to the line that goes out are good observation points, and the hypothesis is tested in this study.

III. Experiment Design

To test our hypothesis, we designed a set of experiments where we varied the *observation points* in a power system where real and reactive power are measured at *from* and *to* buses. We varied those *observation points* based on their LODF values for the line whose outage we wish to detect.

We used three categories of observation points: low LODF, medium LODF, and high LODF. Our experiment design also varied the power flow for the line outage we wish to detect with three categories: low power flow, medium power flow, and high power flow. With the product of three types of observation points and three types of power lines, our experiment design therefore consisted of nine experiments. We randomly repeated each of the nine experiments 24 times to produce an unbiased sample used for statistical inference to validate our conclusions.

Our platform to execute our experiment design consisted of two components:

1) software using MATLAB's MATPOWER package to simulate AC optimal power flow to generate *training* and *testing* data for our classifier
2) software using the `Scikit-learn` Python library for training and testing of a k-nearest neighbor (KNN) classifier

For all of the experiments, we chose to use the *IEEE 118-Bus System*, which represents the American Power System in the Midwestern region of the US during December 1962, for the case studies. This power system has 186 transmission lines, making it difficult to detect when a line goes out when many lines need to be observed. More details on our use of MATPOWER to generate training/testing data is provided in Section III-A.

We used a KNN classifier to detect a line outage using the power data obtained from MATPOWER. Specifically, we trained the KNN classifier with the *training* portion of the data and then evaluated the performance of the KNN classifier using the *testing* portion of the data. We recorded the classification performance measures of precision, recall, and F1-score. More details on our use of the KNN classifier is provided in Section III-B.

Finally, we used inferential statistics to validate our conclusions, we explain in Section III-C.

A. Generating Training and Testing Data

We executed two sets of ACOPF solutions for the *IEEE 118-Bus System*. The first set of solutions represents the power system with all lines functioning, while the second set represents the power system with one line disconnected or *out-of-service*. We gathered information on the real and reactive power flows at the from and to buses of each line in the power system, excluding lines 7, 9, 113, 133, 134, 176, 177, 183, and 184, as their ACOPF solution does not converge when disconnected. As a result, there are 177 lines that can be placed in the *out-of-service* state for detection.

As the load at each bus in real-world power systems fluctuates over time, we simulated realistic data for each line. To achieve this, we randomly varied the power demand for every bus between +/- 1% to +/- 5%. We collected real power at the *from* buses (PF), reactive power at the *from* buses (QF), real power at the *to* buses (PT), and reactive power at the *to* buses (QT) for every line in both OPF solutions. With this data, we created binary class labels where 0 represents a transmission line *in-service* and 1 represents a transmission line *out-of-service*. We gathered 1000 samples for each line: 500 samples for the transmission line *in-service*, and 500 for the transmission line *out-of-service*.

B. KNN Classification

For classification, training data consists of features and the corresponding class labels. In our specific line outage detection classification problem, the features are the PF, QF, PT, and QT measurements, and our class labels are (0: *in-service*) or (1: *out-of-service*). We used the *training* data obtained from the MATPOWER simulations to train a KNN classifier and then used that trained classifier to infer on the *testing* data. We computed performance by checking the inferred labels from the KNN classifier against the ground-truth labels from the MATPOWER simulations. With those comparisons the classification performance measures of precision, recall, and F1-score were computed.

Binary classification produces four outcomes: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). A true positive outcome is when an *out-of-service* line is correctly labeled. A true negative outcome is when an *in-service* line is correctly labeled. A false positive outcome is when an *in-service* line is incorrectly labeled as *out-of-service*. Lastly, a false negative outcome is when an *out-of-service* line is incorrectly labeled as *in-service*.

Precision is the percentage of accurate positive (i.e., *out-of-service*) predictions (i.e., $\frac{TP}{TP+FP}$), recall is the percentage of positive cases correctly predicted (i.e., $\frac{TP}{TP+FN}$), and the F1-score is the harmonic mean of precision and recall.

C. Utilizing Inferential Statistics

We utilized inferential statistics to validate our conclusions regarding line outage classification performance and observation point type based on LODF. Our goal was to have an unbiased sample by selecting random line outage locations, consisting of 24 transmission lines each for low, medium, and high power flows, respectively. We then used these randomly selected line outage locations to test for line outage classification performance at observation points with lines having low, medium, and high LODF values. We collected sample means of precision, recall, and F1-scores, along with 95% confidence intervals computed from the 24 samples.

IV. Experiment Results

To visualize our experiment results, we partitioned the results across three figures (see Figures 2, 3, and 4) each with three subplots. Each of the three figures shows the classification performance for each of the three categories of transmission line for which we detect its service status: Figure 2 for low power flow lines, Figure 3 for medium power flow lines, and Figure 4 for high power flow lines. The three subplots show each of the three performance measures: precision, recall, and F1-score. In each subplot there are three data points, one each for low LODF, medium LODF, and high LODF observation points.

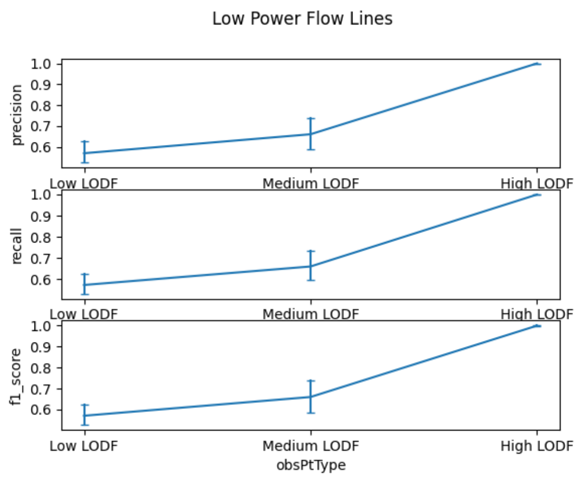

Fig. 2. Detection of low power flow lines with low, medium, and high LODF values.

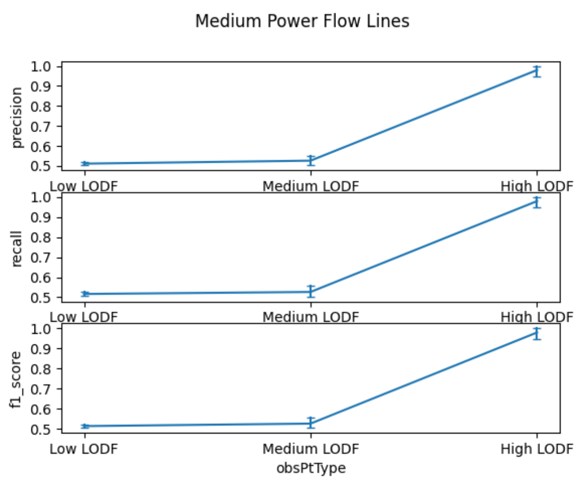

Fig. 3. Detection of medium power flow lines with low, medium, and high LODF values.

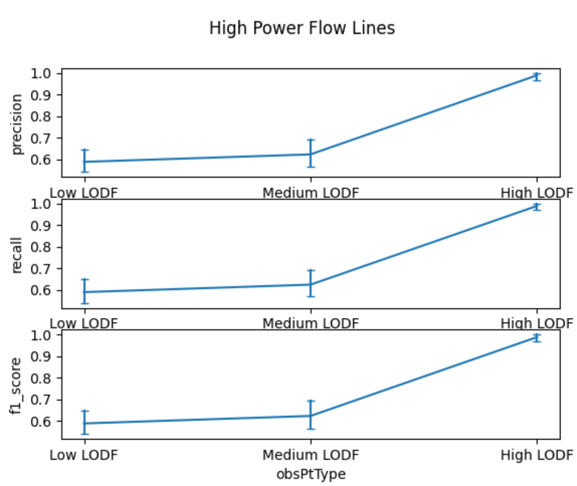

Fig. 4. Detection of high power flow lines with low, medium, and high LODF values.

A. *Low Power Flow Lines*

Figure 2 shows that the best lines to observe when a low power flow line is disconnected are lines with high LODF values. For lines with low LODF values, we get an average precision of 56.94%, an average recall of 57.18%, and an average F1-score of 57.04%. For lines with medium LODF values, we get an average precision of 66.84%, an average recall of 65.93%, and an average F1-score of 65.98%. For lines with high LODF values, we get an average precision of 99.99%, an average recall of 99.97%, and an average F1-score of 99.98%.

B. *Medium Power Flow Lines*

Figure 3 shows that the best lines to observe when a medium power flow line is disconnected are lines with high LODF values. For lines with low LODF values, we get an average precision of 51.17%, an average recall of 51.63%, and an average F1-score of 51.14%. For lines with medium LODF values, we get an average precision of 52.66%, an average recall of 52.26%, and an average F1-score of 52.63%. For lines with high LODF values, we get an average precision of 97.94%, an average recall of 98.05%, and an average F1-score of 97.99%.

C. *High Power Flow Lines*

Figure 4 shows that the best lines to observe when a high power flow line is disconnected are lines with high LODF values. For lines with low LODF values, we get an average precision of 58.93%, an average recall of 58.91%, and an average F1-score of 58.90%. For lines with medium LODF values, we get an average precision of 62.37%, an average recall of 62.38%, and an average F1-score of 62.37%. For lines with high LODF values, we get an average precision of 98.84%, an average recall of 99.01%, and an average F1-score of 98.92%.

D. *Overall Averages for All Lines*

For every line in the power system that can be disconnected, we calculated an average precision, recall, and F1-score for different observation point types. When we use low LODF lines as our observation points, we get an average precision of 55.01%, an average recall of 55.91%, and an average F1-score of 55.69%. When we use medium LODF lines as our observation points, we get an average precision of 60.62%, an average recall of 60.32%, and an average F1-score of 60.33%. When we use high LODF lines as our observation points, we get an average precision of 98.92%, an average recall of 99.01%, and an average F1-score of 98.96%. The results show that no matter what level of power flow that the line disconnected used to carry, lines with high LODF values are in general the best lines to be selected as observation points.

V. Conclusions

In summary, the most effective way to locate line outages is by using observation points with high LODF values corresponding to the line out of service. These points indicate that a significant amount of power from an outage is being distributed to that line, which in turn makes detection much easier using a classification algorithm such as KNN. Conversely, observation points with low or medium LODF values produce relatively poor detection results due to the minimal difference in power flow when a line is disconnected or connected.

Moving forward, our goal is to identify a set of high LODF lines that can be used to effectively monitor as many lines as possible in the system. We have observed that certain line outage locations share the same high LODF lines, which could allow us to group different lines for detection using machine learning and limit the amount of data needed to detect various line outages. Additionally, our line plots have shown a dramatic change from low and medium LODF to high LODF observation points. In the future, we would like to observe the exact point when this change occurs.

References

[1] K. H. LaCommare and J. H. Eto, "Cost of power interruptions to electricity consumers in the United States (US)," *Energy*, vol. 31, no. 12, pp. 1845–1855, Sep. 2006.

[2] Q. Huang, L. Shao and N. Li, "Dynamic Detection of Transmission Line Outages Using Hidden Markov Models," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 2026-2033, May 2016.

[3] M.Z. Zakariya, J. Teh, "A Systematic Review on Cascading Failures Models in Renewable Power Systems with Dynamics Perspective and Protections Modeling," *Electric Power Systems Research*, vol. 214, pp. 108928, 2023.

[4] "2023 State of Reliability Overview," North American Electric Reliability Corporation (NERC), [Online]. Available: https://www.nerc.com/pa/RAPA/PA/Performance%20Analysis%20DL/NERC_SOR_2023_Overview.pdf

[5] H. Zhu and G. B. Giannakis, "Sparse Overcomplete Representations for Efficient Identification of Power Line Outages," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 2215-2224, Nov. 2012.

[6] J. E. Tate and T. J. Overbye, "Line outage detection using phasor angle measurements," *IEEE Transactions on Power Systems*, vol. 23, no. 4, pp. 1644-1652, Nov. 2008.

[7] L. Zhao and W.-Z. Song, "Distributed power-line outage detection based on wide area measurement system," *Sensors*, vol. 14, no. 7, pp. 13114–13133, Jan. 2014.

[8] P. Shah and X. Zhao, "Network Identification Using μ-PMU and Smart Meter Measurements," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 11, pp. 7572-7586, Nov. 2022.

[9] M. Babakmehr, M. G. Simões, M. B. Wakin and F. Harirchi, "Compressive Sensing-Based Topology Identification for Smart Grids," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 2, pp. 532-543, April 2016.

[10] Sadegh Azizi, Mohammad Rezaei Jegarluei, Andreas Martin Kettner, Ahmad Salehi Dobakhshari, "Wide-Area Line Outage Monitoring by Sparse Phasor Measurements," *IEEE Transactions on Power Systems*, vol.38, no.4, pp.3621-3631, 2023.

[11] J. E. Tate and T. J. Overbye, "Double line outage detection using phasor angle measurements," in *Proc. IEEE Power & Energy Soc. General Meeting*, pp. 1-5, 2009.

[12] G. Giannakis, V. Kekatos, N. Gatsis, S. Kim, H. Zhu and B. Wollenberg, "Monitoring and optimization for power grids: A signal processing perspective," *IEEE Signal Processing Magazine*, vol. 30, no. 5, pp. 107-128, May 2013.

[13] J. Wu, J. Xiong and Y. Shi, "Efficient Location Identification of Multiple Line Outages With Limited PMUs in Smart Grids," *IEEE Transactions on Power Systems*, vol. 30, no. 4, pp. 1659-1668, July 2015.

[14] Mehebub Alam, Shubhrajyoti Kundu, Siddhartha Sankar Thakur, Anil Kumar, Sumit Banerje, "A New AC Model for Transmission Line Outage Identification," *Advances in Systems, Control and Automations*, vol. 708, pp. 223, 2021.

[15] Y. Zhao, A. Goldsmith and H. V. Poor, "On PMU location selection for line outage detection in wide-area transmission networks," in *Proc. 2012 IEEE Power and Energy Society General Meeting*, San Diego, CA, USA, 2012, pp. 1-8.

[16] S. Lotfifard, "Sparse Sensing Platform for Line-Outage Identification in Multiarea Power Systems," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 3, pp. 947-955, June 2017.

[17] L. Ma, L. Wang and Z. Liu, "Topology Identification of Distribution Networks Using a Split-EM Based Data-Driven Approach," *IEEE Transactions on Power Systems*, vol. 37, no. 3, pp. 2019-2031, May 2022.

[18] D. Deka, M. Chertkov and S. Backhaus, "Joint Estimation of Topology and Injection Statistics in Distribution Grids With Missing Nodes," *IEEE Transactions on Control of Network Systems*, vol. 7, no. 3, pp. 1391-1403, Sept. 2020.

[19] T. Guler, G. Gross and M. Liu, "Generalized Line Outage Distribution Factors," *IEEE Transactions on Power Systems*, vol. 22, no. 2, pp. 879-881, May 2007.

[20] J. Guo, Y. Fu, Z. Li and M. Shahidehpour, "Direct calculation of line outage distribution factors," *IEEE Transactions on Power Systems*, vol. 24, no. 3, pp. 1633-1634, Aug. 2009.

[21] J. Han, J. Pei, and H. Tong, "Data mining: concepts and techniques". Morgan Kaufmann, 2022.

[22] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, "MATPOWER: Steady-State Operations, Planning and Analysis Tools for Power Systems Research and Education," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, Feb. 2011.