

Impact of Data Sparsity on Machine Learning for Fault Detection in Power System Protection

Julian Oelhaf^{1*}, Georg Kordowich², Changhun Kim¹, Paula Andrea Pérez-Toro¹, Andreas Maier¹
Johann Jäger², Siming Bayer¹

¹Pattern Recognition Lab, Friedrich-Alexander-University Erlangen-Nuremberg

²Institute of Electrical Energy Systems, Friedrich-Alexander-University Erlangen-Nuremberg
Erlangen, Germany

*corresponding author: julian.oelhaf@fau.de

Abstract—Germany’s transition to a renewable energy-based power system is reshaping grid operations, requiring advanced monitoring and control to manage decentralized generation. Machine learning (ML) has emerged as a powerful tool for power system protection, particularly for fault detection (FD) and fault line identification (FLI) in transmission grids. However, ML model reliability depends on data quality and availability. Data sparsity resulting from sensor failures, communication disruptions, or reduced sampling rates poses a challenge to ML-based FD and FLI. Yet, its impact has not been systematically validated prior to this work. In response, we propose a framework to assess the impact of data sparsity on ML-based FD and FLI performance. We simulate realistic data sparsity scenarios, evaluate their impact, derive quantitative insights, and demonstrate the effectiveness of this evaluation strategy by applying it to an existing ML-based framework. Results show the ML model remains robust for FD, maintaining an F1-score of 0.999 ± 0.000 even after a 50x data reduction. In contrast, FLI is more sensitive, with performance decreasing by 55.61% for missing voltage measurements and 9.73% due to communication failures at critical network points. These findings offer actionable insights for optimizing ML models for real-world grid protection. This enables more efficient FD and supports targeted improvements in FLI.

Index Terms—data sparsity, fault detection, machine learning, power grid, sensor reliability

I. INTRODUCTION

Germany’s *Energiewende*, the transition to a renewable energy-based power system, is fundamentally reshaping electric grid operations. The growth of wind, solar, and hydroelectric generation, along with the increase in energy efficiency and demand-side management, presents new challenges for ensuring a secure, stable, and affordable electricity supply [1]. A key aspect of this transition is the expansion and modernization of the power transmission infrastructure, spanning thousands of kilometers, to enable the integration of fluctuating renewable energy while ensuring robust grid operations [2].

As the power system undergoes this transformation, the grid itself must evolve to accommodate *decentralized* energy sources, leading to *dynamic and bidirectional* power flows. Traditional grid control mechanisms, designed for centralized generation, are increasingly insufficient to manage the variability and distributed nature of renewable energy sources [3]. To address these challenges, digitalization has become an essential component of modern grid infrastructure, driving

an unprecedented proliferation of data from a vast network of sensors and measurement devices. These data streams originate from protection relays (PRs) and other monitoring systems, which transfer real-time information to centralized control centers for analysis and decision-making [4]. A key enabler of this transformation is the *IEC 61850* communication standard [5], which enables the integration of remote measurements from protection systems. By improving fault monitoring and enhancing grid resilience, *IEC 61850* plays a crucial role in ensuring efficient and reliable power system operation [6]. However, as the volume and complexity of grid data continue to grow, traditional analytical approaches struggle to extract actionable insights in real-time.

To address this challenge, machine learning (ML) models have emerged as a promising solution for fault detection (FD) [6]–[8], fault line identification (FLI) [9], and protection coordination (PC) [10], enabling more accurate and adaptive fault management while efficiently handling large-scale grid data. However, the performance of ML methods is highly dependent on the quality and availability of training data. An over-reliance on large datasets can introduce vulnerabilities, particularly when these datasets do not adequately reflect real-world data sparsity caused by sensor failures, communication interruptions, and hardware limitations. Such sparsity can compromise ML-based FD and FLI systems, potentially leading to delayed or incorrect protection actions.

To systematically assess the impact of data sparsity, it is essential to consider the factors influencing data availability in power system protection, such as sampling frequency and feature availability. Higher sampling frequencies provide more precise waveform information, but real-world devices often operate at lower rates. Similarly, feature availability, which depends on sensor placement and grid topology, can lead to missing or degraded measurements. In practice, sampling rates typically range from 1 kHz to 10 kHz, varying by relay type and manufacturer. Additionally, legacy infrastructure and communication delays often complicate data availability.

Due to the challenges in obtaining real-world high-resolution fault data, research has relied on simulations. Simulations offer the ability to generate extensive datasets and replicate rare fault scenarios. Many studies [11]–[13] on FD and FLI simulate data sampled at 20 kHz, which can often

be unrealistic in practice. However, the applicability of such studies depends on whether the simulated data truly reflects real-world grid behavior. Addressing the limitations of current ML-based FD and FLI models under data-sparse conditions is crucial for their real-world deployment. A critical gap remains: there is no systematic framework to validate the performance of ML-based protection schemes under varying data quality and availability constraints.

To the best of our knowledge, no prior work has introduced a validation framework to assess the impact of data sparsity on ML-based FD. This study addresses this gap by presenting a proof of concept for evaluating ML models under real-world data constraints with an existing ML model. We systematically analyze how missing sensor data, reduced sampling rates, and temporary communication failures affect ML performance. Focusing on three-phase short circuits within a single power grid topology, we demonstrate the feasibility of a structured validation approach for data quality and availability. Our findings lay the groundwork for assessing the robustness of ML-based FD and FLI methods, providing insights to improve their reliability and real-world deployment in data-limited scenarios.

II. METHODS AND MATERIALS

A. Dataset and Grid Topology

To investigate the impact of data sparsity, we use a widely adopted topology in grid protection studies, as outlined in [14]. The power grid simulation is based on the “Double Line” topology, shown in Fig. 1, and generated using DIGSILENT’s PowerFactory¹ software, with simulations conducted by an expert in electrical power engineering. This framework extends the methodology from [15]. Key parameters, including external grid settings, load conditions, transmission line lengths, and short circuit locations, were systematically varied to simulate a broad range of fault scenarios. The rationale behind this parameter variation, known as domain randomization, is to improve the robustness of ML training and facilitate the transition from simulation to real-world scenarios. The selected parameters aim to accurately represent realistic grid conditions while covering a wide range of possible scenarios. The parameter ranges were based on the works by [16] and [17]. For further details on the parameter randomization process, refer to our previous study [8]. This study specifically focuses on three-phase faults, which, although accounting for only 5% of fault occurrences [18], are responsible for the most severe disruptions in the grid. The objective is to assess whether these faults can be reliably detected despite the challenges posed by data sparsity.

Each sample has a unique fault configuration, lasting 1 s at a 20 kHz sampling rate. Since PRs installed at Buses measure three-phase current and voltage ($I_A, I_B, I_C, V_A, V_B, V_C$), each device records a total of six signals. Given that eight PRs are deployed on the transmission lines, the system generates 48 measurements per time step, serving as input for FD and FLI.

In total, the network consists of three Buses (Bus 1–3) connected by four transmission lines. Bus 1 links to an external grid, while Bus 2 acts as a junction with four PRs monitoring two incoming and two outgoing lines. Bus 3 is connected to two transmission lines with PRs. In this study, we focus on the protection of transmission lines.

Current and voltage signals at each PR are continuously monitored for real-time protection and FD. These measurements form a multivariate time series essential for FD and FLI, with the current and voltage signals for each phase given by:

$$I_{PR}(t) = (I_A(t), I_B(t), I_C(t)), \quad t \in [0, 1] \text{ s} \quad (1)$$

$$V_{PR}(t) = (V_A(t), V_B(t), V_C(t)), \quad t \in [0, 1] \text{ s} \quad (2)$$

where subscripts A, B, C correspond to the three phases. The combined multivariate time series per PR is then:

$$X_{PR}(t) = \begin{bmatrix} I_{PR}(t) \\ V_{PR}(t) \end{bmatrix}, \quad t \in [0, 1] \text{ s} \quad (3)$$

B. Data Preprocessing

We preprocess the raw simulation data by trimming each sample to ± 80 ms around the fault start to capture critical events before and after the fault. To simulate real-time constraints, we slide the time series with a 5 ms step, ensuring overlap between snippets. We evaluate the impact of different window lengths: 10 ms, 20 ms, 30 ms, 40 ms, and 50 ms. These window lengths were chosen to balance between capturing enough information for accurate FD while adhering to the practical constraints of real-time processing. Additionally, with a 50 Hz power system, one period corresponds to a duration of 20 ms.

C. Machine Learning Models for Fault Detection and Line Identification

FD is formulated as a binary classification task, while FLI as a four-class classification problem. The feature space consists of concatenated voltage and current measurements from the four transmission lines, forming a multivariate time series over a predefined window length. Each window is labeled based on the presence or absence of a three-phase short-circuit fault, and, where applicable, the corresponding faulted transmission line, as described in [8].

For consistency across all tasks, the same ML model is used in all experiments. The Random Forest (RF) classifier is selected based on our comparative study [8], which demonstrated its favorable trade-off between performance and runtime. The reported F1-score is 0.998 for FD, 0.985 for FLI, with a runtime of 16.8 ms.

To ensure a thorough analysis of the impact of data sparsity, 4000 samples are simulated using the procedure outlined in Sec. II-A. All experiments follow a 5-fold cross-validation scheme, ensuring consistent evaluation on unseen data. The F1-score is used as the evaluation metric due to its balanced consideration of both sensitivity and specificity. Model implementation, training, cross-validation, and evaluation are carried out using *scikit-learn* [19].

¹<https://www.digsilent.de/en/>

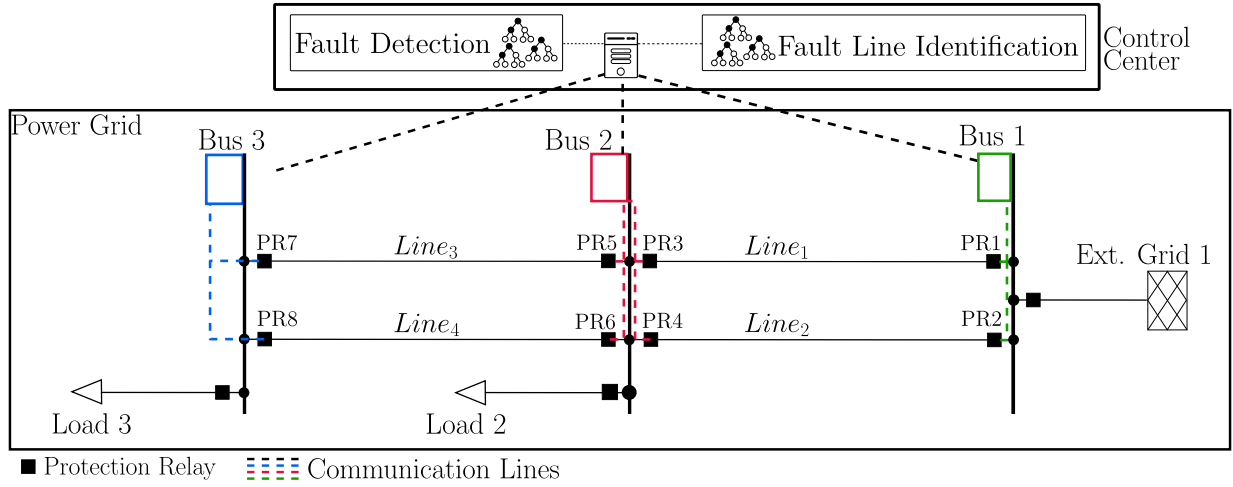


Fig. 1. Schematic of the “Double Line” grid topology. The diagram illustrates the transmission lines, PRs, and communication devices. Measurements are recorded at the PRs and transmitted to the substations at each bus. From there, the data is relayed to the control center for FD and FLI.

D. Scenarios of Data Sparsity

This study examines the impact of data sparsity on FD and FLI in power systems by simulating realistic cases of missing or incomplete data. In collaboration with an expert in electrical power engineering, we identified plausible scenarios that may occur in a power grid, including hardware failures, sensor malfunctions, and communication disruptions. Additionally, we account for legacy infrastructure operating at lower sampling rates. Each scenario is described in detail below.

Missing Voltage or Current Data: This scenario simulates sensor failures by setting either voltage or current measurements to zero while keeping the remaining data unchanged. Such failures can result from sensor malfunctions or communication issues.

Reduced Temporal Resolution: This scenario models the effect of data being collected at a lower sampling rate due to hardware limitations. To simulate this, data points are discarded based on a predefined downsampling factor. This technique also accounts for real-world constraints, such as limited communication bandwidth, which can restrict the volume of data transmitted over time. We reduce the original sampling frequency to 10 kHz, 4 kHz, 2 kHz, 800 Hz, and 400 Hz.

Component Failures: This scenario is intended to model the failure of critical system components by removing all measurements from the affected component. The loss of communication with a substation would result in the loss of measurements from all PRs connected to the corresponding Bus. We model the failure for Buses 1, 2, and 3. Furthermore, we model the individual failure of each of the eight PRs. Buses and PRs are shown in Fig. 1.

Phase-Specific Data Loss: This scenario is designed to mimic the loss of measurements from one of the phases in a three-phase system. Such a loss can be attributed to physical failures, such as damaged sensors or transmission lines, or to disruptions in communication. In this case, one of the three

phases, Phase A, Phase B, or Phase C, experiences an outage, resulting in missing measurements.

Communication Loss: This scenario simulates a complete loss of communication due to issues at the receiver, resulting in temporal data transfer disruptions over specific time intervals. During these periods, measurements are missing. We simulated communication outages ranging from 5 ms to 45 ms, in steps of 5 ms. Only scenarios where the communication loss period is shorter than the window length are considered.

TABLE I
OVERVIEW OF DATA SPARSITY SCENARIOS AND PARAMETER VARIATIONS.

Sparsity Scenario	Parameter Values
Missing Voltage	True, False
Missing Current	True, False
Reduced Sampling Frequency	10 kHz, 4 kHz, 2 kHz 800 Hz, 400 Hz
Substation Communication Failure	1, 2, 3
Relay Communication Failure	1 - 8
Phase Measurement Failure	A, B, C
Temporal Communication Loss	5 ms - 45 ms

Tab. I summarizes the different data sparsity scenarios and their corresponding parameter variations. Each scenario introduces a unique challenge for FD models, helping us evaluate their robustness under real-world conditions. By incorporating these variations into our study, we systematically assess how different types of data sparsity can affect FD performance. This evaluation is crucial for understanding the resilience of ML models in real-world applications, where missing or degraded data is a common challenge.

III. EXPERIMENTS AND RESULTS

In this study, we perform experiments on two tasks: FD and FLI. For each task, we evaluate five different window lengths and five data sparsity scenarios, as introduced in Sec. II-D, with the parameter values summarized in Tab. I. This results in a total of 270 experiment combinations.

For all scenarios, the F1-score of the RF model across the five window lengths is used as the baseline: 0.998 ± 0.001 for FD and 0.997 ± 0.001 for FLI. However, it is important to note that, due to the differing number of available window samples—since only fault windows are used for the FLI task—and the fact that FD is a binary classification task while FLI is a multi-class classification task, the two tasks cannot be directly compared.

Missing Voltage or Current Data: The results presented in Tab. II show the impact of missing V and I measurements on model performance. For the FD task, the model performed slightly better with missing V measurements (+0.02%) but showed a minor decrease in performance when I measurements were missing (−0.16%). In contrast, for the FLI task, the model experienced a significant performance drop when V measurements were missing (−55.61%), while the removal of I measurements had a negligible impact (+0.04%).

TABLE II

IMPACT OF LOSS OF VOLTAGE AND CURRENT MEASUREMENTS ON MODEL PERFORMANCE. THE PERFORMANCE DROP IS MEASURED RELATIVE TO THE NO-FAILURE CASE.

Task	Measurement	F1↑	Change (%)
FD	Missing V	1.000 ± 0.000	+0.02%
	Missing I	0.998 ± 0.001	−0.16%
FLI	Missing V	0.443 ± 0.008	−55.61%
	Missing I	0.998 ± 0.001	0.04%

Reduced Temporal Resolution: Lower sampling rates had minimal impact on the performance, with slight variations across frequencies. For both FD and FLI tasks, performance remained nearly unchanged even at 400 Hz, a 50× reduction from the baseline. This indicates that the model is quite robust to lower sampling rates, suggesting its capability to generalize well even when the input data is significantly compressed. The F1-score showed negligible improvements between 0.01% and 0.09%. The largest change occurred in the FLI task at 400 Hz, reaching an F1-score of 0.998 (+0.09%). Due to space constraints, detailed results are omitted for this scenario.

Component Failures: A notable factor influencing performance was the scenario involving Bus component failures. While the impact on the FD task was relatively minimal compared to the no-failure case, the FLI task exhibited a more pronounced performance degradation. Specifically, the failure of Bus 3 resulted in a minor drop of −2.85%, while Bus 2 led to a more substantial decline of −9.73%. However, Bus 1 failure caused a slight reduction of only −0.28% (see Tab. III).

Phase-Specific Data Loss: The absence of phase measurements had a negligible effect on the model’s performance, with only minor fluctuations across the different phases. For both FD and FLI tasks, the model exhibited nearly identical performance to the baseline, even when data from individual phases were missing. In the case of the FD task, slight increases in the F1-score were observed for phases A and C (+0.01%), while phase B showed no change compared to the baseline. For the FLI task, phase A demonstrated a small

TABLE III

IMPACT OF BUS FAILURES ON MODEL PERFORMANCE. THE PERFORMANCE DROP IS MEASURED RELATIVE TO THE NO-FAILURE CASE.

Task	Bus Nr.	F1↑	Change (%)
FD	Bus 1	1.000 ± 0.000	+0.02%
	Bus 2	0.999 ± 0.000	0.00%
	Bus 3	0.999 ± 0.000	0.00%
FLI	Bus 1	0.995 ± 0.001	−0.28%
	Bus 2	0.900 ± 0.003	−9.73%
	Bus 3	0.969 ± 0.002	−2.85%

improvement of +0.04%, whereas phases B and C exhibited a minimal rise of +0.01%.

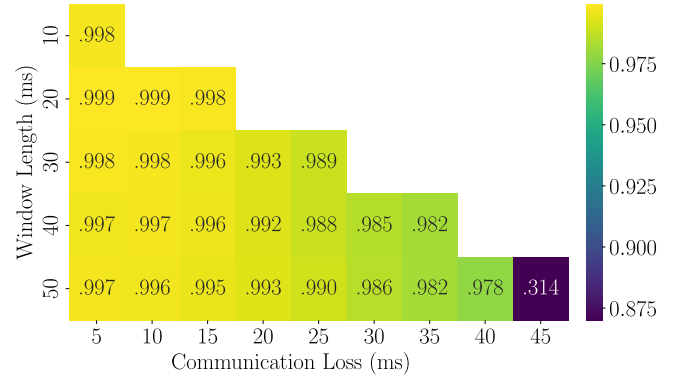


Fig. 2. Heatmap of the Fault Line Identification task F1-scores under cases of temporal communication loss.

Communication Loss: Short communication losses had a negligible effect on the model’s FD performance. For outages of up to 20 ms, the F1-score remained nearly unchanged, with a maximum drop of only −0.13%. As the duration of the outage increased, the performance gradually decreased, reaching a drop of −0.90% at 40 ms. However, a 45 ms communication loss caused a substantial drop in F1-score by 32.8%, reducing it to 0.672.

The FLI task was more sensitive to temporary communication losses, as shown in Fig. 2. Even a brief outage of 30 ms led to a performance decline of −1.20%, reducing the F1-score to 0.985. For a 45 ms outage, performance dropped sharply to 0.314, corresponding to a decrease of −68.6%. In both tasks, a communication loss of 45 ms corresponds to an approximate information loss of 90%, severely limiting the model’s ability to make accurate predictions.

IV. DISCUSSION

The results indicate that the ML model remains robust against most constraints and handles data sparsity well. Model performance remained stable across all window lengths, as indicated by the low standard deviation, with FD proving particularly resilient. For both tasks, the model handled missing I measurements well. However, FLI was sensitive to missing V measurements, highlighting their importance for identifying the faulted line. Reducing the temporal resolution had minimal impact, as the RF classifier maintained stable

performance even at 400 Hz. This suggests that lower sampling frequencies are sufficient to detect and identify the faulted line in three-phase short circuits. Component failures had little overall effect, with FD unaffected. The negligible impact of communication failure at individual PRs likely result from each line having two PRs, which compensate each other. However, the results reveal that measurements from Bus 2 are critical for FLI, as performance dropped significantly when these measurements were unavailable. Phase-specific data losses had minimal impact. This may be due to the symmetrical nature of three-phase faults, reducing the dependency on individual phase measurements for FD and FLI. For temporary communication losses, FD remained robust, but FLI declined with longer data loss durations. A 5 ms loss in the 10 ms window had less impact than a 25 ms loss in the 50 ms window, despite both being 50% of data loss.

These findings highlight the RF model's resilience under most data constraints, particularly for FD, while emphasizing the greater sensitivity of FLI to missing voltage measurements, Bus 2 failures, and extended communication losses. The results indicate that FD requires significantly less data than previously assumed. This could allow FD to be performed more efficiently, freeing up computational resources and time for more complex FLI tasks.

V. CONCLUSION

This study demonstrates the effectiveness of a systematic evaluation strategy for assessing the impact of data sparsity on the performance of ML-based FD and FLI. This approach addresses a critical gap in current research and offers insights that could guide the development of more robust ML applications in this field. Furthermore, it offers comprehensive quantitative findings using an existing ML method, focusing on three-phase short circuits in transmission lines, resulting in a proof of concept for evaluating ML models under real-world data constraints. Our quantitative results show that the ML model for FD remains robust against sparse and missing data, while the model for FLI is more sensitive to data sparsity. The findings suggest that lower data resolutions can yield faster and equally accurate FD results. Future research should expand this validation framework to assess how data sparsity affects various ML approaches and apply it to different grid topologies. Additionally, the framework should be extended to include additional fault types (e.g., two-phase, ground faults), load conditions, multiple fault scenarios, and resilience to noisy data. This would enable more comprehensive, reliable, and real-world applicable validation of ML models for power system protection.

ACKNOWLEDGMENT

This project was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 535389056.

REFERENCES

- [1] M. J. B. Kabeyi *et al.*, "Sustainable Energy Transition for Renewable and Low Carbon Grid Electricity Generation and Supply," *Frontiers in Energy Research*, vol. 9, p. 743114, Mar. 2022. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fenrg.2021.743114/full>
- [2] B. für Wirtschaft und Klimaschutz, "Aktueller Stand des Netzausbau (Übertragungsnetz)," https://www.bmwk.de/Redaktion/DE/Downloads/M-O/netzausbau-schreitet-voran.pdf?__blob=publicationFile&v=17, 2024, accessed: 2025-03-11.
- [3] Central Power Research Institute, V. Telukunta *et al.*, "Protection challenges under bulk penetration of renewable energy resources in power systems: A review," *CSEE Journal of Power and Energy Systems*, vol. 3, no. 4, pp. 365–379, Dec. 2017. [Online]. Available: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8233582>
- [4] M. A. Aftab *et al.*, "IEC 61850 based substation automation system: A survey," *International Journal of Electrical Power & Energy Systems*, vol. 120, p. 106008, Sep. 2020. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0142061520304646>
- [5] I. E. Commission, "IEC 61850:2024 SER | IEC," May 2024. [Online]. Available: <https://webstore.iec.ch/en/publication/6028>
- [6] S. Suhail Hussain *et al.*, "A novel PRP based deterministic, redundant and resilient IEC 61850 substation communication architecture," *Perspectives in Science*, vol. 8, pp. 747–750, Sep. 2016. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S2213020916302166>
- [7] A. Abdullah, "Ultrafast Transmission Line Fault Detection Using a DWT-Based ANN," *IEEE Transactions on Industry Applications*, vol. 54, no. 2, pp. 1182–1193, Mar. 2018. [Online]. Available: <http://ieeexplore.ieee.org/document/8110712/>
- [8] J. Oelhaf *et al.*, "A Systematic Evaluation of Machine Learning Methods for Fault Detection and Line Identification in Electrical Power Grids," in *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025, pp. 1–5.
- [9] A. Zainab *et al.*, "Faulted Line Identification and Localization in Power System using Machine Learning Techniques," in *2019 IEEE International Conference on Big Data (Big Data)*. Los Angeles, CA, USA: IEEE, Dec. 2019, pp. 2975–2981. [Online]. Available: <https://ieeexplore.ieee.org/document/9006377/>
- [10] G. Kordowich *et al.*, "A hybrid Protection Scheme based on Deep Reinforcement Learning," in *2022 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*. Novi Sad, Serbia: IEEE, Oct. 2022, pp. 1–6. [Online]. Available: <https://ieeexplore.ieee.org/document/9960539/>
- [11] M. Chingshom *et al.*, "Fault Detection and Classification in Smart Grid Using Machine Learning Approach," in *2024 International Conference on Advancement in Renewable Energy and Intelligent Systems (AREIS)*. Thrissur, India: IEEE, Dec. 2024, pp. 1–6. [Online]. Available: <https://ieeexplore.ieee.org/document/10893627/>
- [12] A. A. Jabbar *et al.*, "Power Quality Analytical Methodology Using Signal Processing and Machine Learning Techniques for Grid Fault Classification and Location Detection," in *2024 IEEE 12th International Conference on Smart Energy Grid Engineering (SEGE)*. Oshawa, ON, Canada: IEEE, Aug. 2024, pp. 52–56. [Online]. Available: <https://ieeexplore.ieee.org/document/10739482/>
- [13] M. Najafzadeh *et al.*, "Fault Detection, Classification and Localization Along the Power Grid Line Using Optimized Machine Learning Algorithms," *International Journal of Computational Intelligence Systems*, vol. 17, no. 1, p. 49, Mar. 2024. [Online]. Available: <https://link.springer.com/10.1007/s44196-024-00434-7>
- [14] G. J. Meyer, *et al.*, "Hybrid fuzzy evaluation algorithm for power system protection security assessment," *Electric Power Systems Research*, vol. 189, p. 106555, Dec. 2020. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S037877962030359X>
- [15] M. Wang *et al.*, "A Generic Data Generation Framework for Short Circuit Detection Training of Neural Networks," in *PESS + PELSS 2022; Power and Energy Student Summit*. Kassel, Germany: VDE, Nov. 2022, pp. 49–54. [Online]. Available: <https://ieeexplore.ieee.org/document/10104226>
- [16] R. Roeper and Mitlehner, Friedrich, *Kurzschlußströme in Drehstromnetzen*, 6th ed. Publicis Corporate Publishing, 1984.
- [17] D. Oeding and B. R. Oswald, *Elektrische Kraftwerke und Netze*, 8th ed. Springer Berlin Heidelberg, 2016. [Online]. Available: <http://link.springer.com/10.1007/978-3-662-52703-0>
- [18] T. Gonen, *et al.*, *Electric Power Distribution Engineering*, 3rd Edition, 3rd ed. CRC Press, 2015, oCLC: 1105773230.
- [19] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.