Learning from Future: Prediction-based Data Augmentation to Enhance Power Grids Fault Detection

Jennifer Rogers, William Danilczyk, Hui Lin, Yan (Lindsay) Sun
Department of Electrical, Computer and Biomedical Engineering
University of Rhode Island
Kingston RI, United States
{jarogers, wdanilczyk, huilin, yansun}@uri.edu

***Abstract*—As power grids modernize to include wide area monitoring and advanced metering infrastructures, the applications for data-driven methods based on artificial intelligence (AI) for situational awareness, reliability, and security continue to grow. However, obtaining effective training data that captures normal and anomalous operations in power systems remains a great challenge. This paper proposes a data-augmentation method that can efficiently build effective training data sets, benefiting various data-driven applications used in power grids without increasing the size of training data set. By leveraging load prediction in smart grids, we obtain the knowledge of future operating conditions and potential anomalies, which are integrated with the historical data in the training data set. By utilizing this augmented training data set, we significantly increase the accuracy of data-driven fault detection, e.g., 8.6% on average, compared to the results trained based on historical data only.**

***Index Terms*—Data augmentation, machine learning, fault detection.**

I. Introduction

The reliability and security of power grids are essential in ensuring economic sectors' continuous operations and maintaining public safety [1]. The advent of smart grids and advanced metering infrastructure provides new opportunities for anomaly detection, which can be widely applied to various problems, including energy fraud detection [2], demand management [3], and fault detection [4]. Among these applications, accurately detecting and locating faults in power systems is critical to shorten response time and prevent cascading outages, ensuring reliability and stability of the grid [5].

In a conventional power system, model-based methods dominate fault detection and location, mainly by solving explicit physical models of power grids. For example, studies in [6], [7] use the model of state estimation to identify different types of faults in transmission and distribution systems. These methods can incur long latency by processing sufficient measurements in grids' physical model to ensure accurate detection. With the increasing adoption of renewable energy and dynamic configuration, model-based methods can struggle to capture all variables and dependencies in power systems, making them infeasible in certain situations.

978-1-6654-9921-7/22/$31.00 ©2022 IEEE

Alternative to the model-based methods, many studies have exploited the data collected from wide area monitoring (WAM) systems to implement data-driven methods [5]. These methods are model-free and can detect faults or anomalies with a small latency by making inferences in a pre-trained deep neural network [8], [9]. Furthermore, studies show that many power grid applications beyond fault detection, including demand responses and intrusion detection, can benefit from data-driven approaches if their models are trained properly [10].

While data driven methods have become the most prevalent forms of fault detection in the current research, there is still a wide research gap affecting their performance, i.e., how to obtain effective training data. Solving this problem by exclusively using historical data is challenging for two reasons. *First*, slow mechanical inertia in power grids requires a long history to collect various operating conditions. *Second*, some studies leverage Monte Carlo simulation, which can only produce random operating conditions. Those historical or randomly-generated conditions can fail to characterize the future operating conditions, under which faults occur, significantly downgrading the performance of data-driven approaches.

To fill this gap, we propose an original data-augmentation method that can efficiently build effective training data sets. Current data-augmentation methods used in image processing enhance training data sets by manipulating existing data and increasing the data set size. These methods can become inefficient and ineffective for power grids' applications, because (i) creating various operating conditions require heavy computation, (ii) a large training data set can inevitably increase the training overhead, and (iii) newly added training data may not fully capture the characteristics of future operations affected by unexpected factors, e.g., weather or human involvements.

Instead of randomly adding training data, our original data-augmentation method leverages load prediction to add a small amount of training data closely related to upcoming operating conditions and potential faults. Load prediction is commonly used to optimize power grid stability and economical benefits; to our knowledge, this is the first work to integrate load prediction with data-augmentation and improve the performance of data-driven applications in power grids. To further reduce training overhead, we introduce two design variations that synthesize predicted conditions with historical ones and seamlessly apply the augmented data set into existing training procedure. Evaluations of two IEEE test systems, simulated in two different environments, show a significant improvement in data-driven fault detection, e.g., approximately 8.6% in detection accuracy on average, even without increasing the size of the training data set.

Even though we focus on one type of fault detection in this paper, the proposed data-augmentation methods can be integrated with other data-driven applications, which are gradually equipped in modern power grids. As data driven approaches offer a critical alternative to traditional model-based methods, we believe that the data-augmentation methods can benefit a wide range of applications, which we leave as future work.

The rest of the paper is organized as follows. Section II introduces related research in the areas of fault detection, data augmentation and load prediction. Section III describes the proposed data augmentation and fault detection methods, Section IV details the experimentation and results, and conclusions are drawn in Section V.

II. Related Works

Data Driven Fault Detection. Detection of faults in the smart grid is critical to shorten response time and prevent cascading outages. In [5] the authors implement data driven methods such as pursuit decomposition, hidden Markov models and k-means clustering to detect, identify, and locate faults. Other machine learning techniques have been employed for fault prediction in [11], where the authors implement long short term memory (LSTM) and support vector machines (SVM) to predict the occurrence of faults based on historical information. Most recently, research has focused on deep learning through artificial neural networks (ANN). Specifically convolutional neural networks (CNNs), which leverage convolution layers to catch spatial inter-dependency of the input data, demonstrate big success in areas like object identification and image segmentation [12]. In smart grids, CNNs have been employed for applications such as energy theft detection [13], false data injection detection [14] and fault diagnosis [15]. Our previous work demonstrated a CNN could be employed to locate bus faults with a high degree of accuracy for a single operating condition [16]. *The focus of this work is not to propose a new fault detection method; we will improve the performance of the previous anomaly detection methods by augmenting the training data with future operating conditions.*

Data Augmentation. While there remains a research gap in data augmentation for smart grid applications, it is applied to various other disciplines such as image processing, medical research, and computer networks [17], [18]. In image processing, data engineers use manual manipulation to create more images, through rotating or mirroring. More complicated approaches, such as generative adversarial networks (GAN), use the statistical qualities of existing data. Specifically, many studies use GAN to create data that can better train anomaly detection tools in different applications, e.g., detecting cardiac dysfunction in ECG signals [18]. *Our data-augmentation method distinguishes itself from previous work by creating data that follows the physical laws and matches statistical features of historical data in power grids.* This is achieved by predicting the future operating states of the power system.

Load Prediction. Power grid operations vary with time, continuously accommodating load demands from residential and industrial sectors. Load prediction plays a critical role in estimating the trajectory of load demands in the future by integrating various data, including statistical characteristics of historical load demands, weather conditions, and user patterns. Long-term prediction is helpful for power system infrastructure planning, while short-term forecasting is widely used to adjust runtime system operations [19]. Hernandez et al. provide a comprehensive review of load prediction algorithms in the last 2+ decades; most recent research has focused on using ANN [20] and recurrent neural networks, e.g., LSTM [21], [22].

III. Prediction Based Data Augmentation

A. Data-driven Fault Detection Overview

This paper uses the data-driven fault detection presented in [16] as an application to demonstrate the advancement of the proposed data-augmentation method. Our previous method detects faults by leveraging a CNN to process data collected from increasingly deployed smart meters. Our CNN model involves two feature layers, each of which included three convolutional layers in sequence. Each convolutional layer had a kernel size of three, a stride of one, and a single zero padding.

The CNN model takes a set of time-series data as inputs and infers the presence of faults. Specifically, each input data point is represented by time-dependent three-dimensional tensor A_t. An entry $A_t[i, j, k]$ specifies the voltage phasors at phase k of bus j. The value i indicates the time stamp at $t + i$ when a meter, like a phasor measurement unit (PMU), samples the corresponding value. Consequently, A_t includes time-series voltage phasors sampled at high frequency within a small period starting at t. To feed various data points to the CNN model, we collect A_t at different times t, when a power grid experiences different operating conditions. For example, in Fig. 1(a), existing data-driven methods collect data points from history only, e.g., collecting A_t with $t_0 \leq t \leq t_1$. After collecting historical data, training starts at t_1 and ends at t_2, taking δ time units.

Assuming the training period is short and a fault happens after the training, i.e., $t_a > t_2$, the accuracy of machine learning models heavily relies on the training. If the training data set fails to include the characteristics or probability distributions in future events, data-driven models tend to make incorrect inferences. Consequently, we encounter significant research gaps to ensure the performance of deep-learning models in smart grids. ***First***, many unexpected factors, e.g., weather or human involvements, are not reflected in the historical data. ***Second***, even if we build training data from a long history or simulation, the randomly-generated operating conditions may not be representative of future conditions during which faults occur. To make things worse, blindly enlarging the training data set also increases the training period δ; faults becomes more probable to occur prior to completing the training and force the inference to be made based on the subset of training data. Consequently, building effective training data requires we add operating conditions consistent with future conditions without significantly increasing the training overhead.

B. Design of Prediction-based Data Augmentation

To bridge the gaps in the existing training procedure, we propose a data-augmentation method that enriches the training data set with the predicted knowledge of future operating conditions in a power grid. In Fig. 1, we present this idea in a timeline, demonstrating the benefits and its relationship with the existing training procedure.

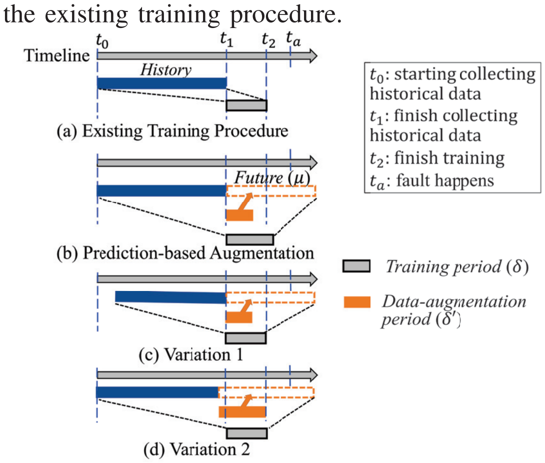


Fig. 1: Prediction-based training: augmented historical data with the knowledge relevant to future and anomaly operations.

As shown in Fig. 1(b), the highlighted orange bar specifies the time period (marked as δ') that it takes to augment the training data, creating predicted data for the future time period (marked as μ) and simulating faults that occurred during historical and future operations. Consequently, the training uses both historical and augmented data, including operations closely relevant to faults.

Prediction-based data augmentation can be seamlessly integrated with existing training procedures with a small overhead. Simulating faults on historical operations introduces inconsequential overheads, because it can be performed while historical data is collected. Creating predicted operations and subsequently simulating faults (e.g., during period μ) can be performed in parallel while training the model with the historical data. For example, predicting the operation 20 hours ahead in a high-fidelity environment such as OPAL-RT (i.e., $\mu \approx 20$ hours) based on historical data and creating faults for the 20-hour predicted operations takes around 1.5 hours in total (i.e., $\delta' \approx 1.5$). This augmentation period can be completely overlapped with the training of historical data, which lasts 3.5 hours. However, training the model with the augmented data adds another 0.5 hours, which makes $\delta \approx 3.5+0.5 = 4$ hours; the 0.5-hour of overhead is tolerable in our experiments.

The predicted conditions can track the significant changes of grid operations, complementing missing knowledge in historical data. Instead of predicting a few specific faults, we enhance the training data with diverse knowledge that can be relevant to fault conditions. Even if the prediction does not have high accuracy, the enhanced training data can still increase the fault detection accuracy (see evaluation in Section IV). This work focuses on supervised learning; we believe that prediction-based augmentation can also help other data-driven methods, e.g., reinforcement learning, which will be our future work.

Design Variations to Remedy Data-augmentation Overheads. Based on the proposed design, we expect to encounter two types of overheads: (i) additional training on the predicted and fault conditions, and (ii) additional latency caused by a too-long data-augmentation period when $\delta' > \delta$. We propose two design variations to remedy the overheads. When the size of a training data set becomes too large, we propose variation 1 in Fig. 1(c), which reduces the amount of historical data used in the training data set. For example, in our evaluation, when we apply this variation, the data-augmentation can still improve the fault detection performance (see Section IV for details). When additional latency, δ', is too large (i.e., $\delta^{'} > \delta$), we propose variation 2 in Fig. 1(d). Specifically, we stop collecting historical data earlier (e.g., at $t_1 - (\delta^{'} - \delta)$) so that we can start data-augmentation earlier, compensating for the additional latency.

C. Components of Prediction-based Data Augmentation

In Fig. 2, we present the components of prediction-based data-augmentation, whose details are presented in the following paragraphs.

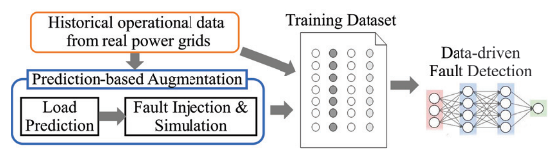


Fig. 2: Components of the prediction-based data-augmentation.

Load Prediction. In this paper, we use LSTM as the method to predict transactions in power grids, benefiting from its good performance in processing time-series data in applications like natural language processing. However, the proposed data-augmentation method is not restricted by the specific load-prediction algorithm. Other load predictions used in actual utility companies can be easily integrated.

After estimating load demands, we use deterministic algorithms, e.g., optimal power flow analysis or economic dispatch, to determine remaining unknown operating parameters, e.g., generation outputs [23]. The resultant predicted and benign operating conditions are added to the training data.

Fault Injection & Simulation. In addition to enhancing the training data with benign predicted operations, we further simulate faults on the predicted and historical operations. The resultant simulated fault conditions are added to the training data as well, which can better balance the training data by including the data labeled with both considered categories (i.e., “normal” and “anomaly”). Similar to the load prediction component, the fault injection and simulation component can benefit from high-order contingency screening implemented in actual utility environments.

The implementations of load prediction and fault injection/simulation components *are not restricted by a specific simulation environment*. This work achieves the prediction-based data augmentation by advanced functionalities in an intelligent digital twin. A traditional digital twin uses a high-fidelity simulation of cyber and physical infrastructures to process runtime measurements from a real environment. Its capability to parallelize the computation of real-time data enables system administrators to optimize grid operations. Our recent work developed an advanced intelligent digital twin by including machine learning techniques to quantify patterns of external factors, e.g., user behavior and weather changes [23]. *Developing a new digital twin is beyond the scope of this work*. Instead, this work uses the implementation in [23] to create predicted data and simulate faults to enrich the existing training data set by retrieving historical measurements.

IV. Evaluation

A. Experiment Setup

We conduct evaluations in two power systems, i.e., IEEE 9-bus and 39-bus systems. To simulate the normal variation of operations in the power grid, we built a profile representing the changes of load demands based on a public data set, *ACTIVSg2000*s [24]. This data set includes 8,784 different operating conditions, based on hourly load demands observed at real substations for one year. We built six evaluation cases; each case included 100 hours of those operating conditions which were randomly selected. In Table I, we list the starting and ending timestamps in the *ACTIVSg2000* data set for each evaluation case; the first 80 hours of operations are fully or partially used as historical data, while the remaining 20 hours of operations are always used as the testing data set.

TABLE I: Evaluation cases.

Case	Starting/Ending Time	Case	Starting/Ending Time
Case 1	0-100 hours	***Case 4***	4000-4100 hours
Case 2	500-600 hours	***Case 5***	7000-7100 hours
Case 3	1900-2000 hours	***Case 6***	8684-8784 hours

For each evaluation case, we performed experiments under four *training scenarios*. All four scenarios used the same testing data set; they differ in the composition of the training data set, as shown in Table II. The first, ***Base-60*** scenario, used 60 hours of historical operating conditions in the training data set. In the second, ***Augmented*** scenario, we combined the 60-hour historical and the 20-hour predicted conditions. The third, ***V1-Augmented*** scenario (corresponding to the variation 1 design in Section III.B), included only 40 hours of historical conditions and the 20 hours of predicted conditions in the training data set. In this scenario, the training data set has the same size as the training data set used in the Base-60 scenario. In addition, we compare all of these results to the training scenario that used the 80 historical hours of operating conditions (labeled as ***Base-80***).

Discussion. Note that we used the subset of the 80-hour of historical conditions in the first three training scenarios

TABLE II: Training scenarios.

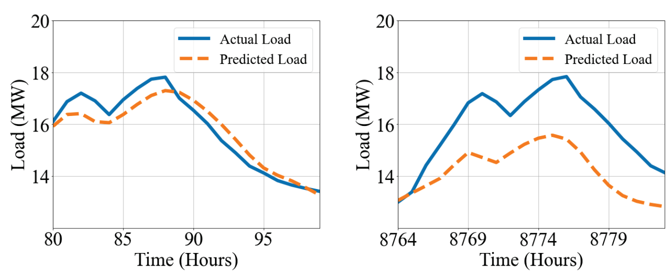
Scenario	Composition of Training Data
Base-60	60 hrs historical data
Augmented	60 hrs historical data + 20 hrs predicted data
V1-Augmented	40 hrs historical data + 20 hrs predicted data
Base-80	80 hrs historical data

to represent the situation when there are not enough training data. Our evaluations show that data-augmentation can significantly improve the fault detection accuracy in those specific situations, where the performance of data-driven applications usually downgrade.

Our implementation includes variation 1 of the prediction-based data augmentation (see Fig. 1), but not variation 2. This is because the period of data augmentation is much smaller than the training period (i.e., $\delta' < \delta$) in our experiments. Note that even if we intentionally delay the data-augmentation, variation 2 of the data augmentation would present same-level of improvements in the fault detection as shown in the V1-Augmented scenario. In future work, we will further explore data-driven approaches in other critical infrastructures, in which data-augmentation may have much longer latency.

B. Prediction-based Data Augmentation Implementation

Load Prediction Component. We leveraged the LSTM model to implement the load prediction component. Specifically, the LSTM model included a linear prediction model, using mean squared error (MSE) as the loss function and the Adam optimization function to train the model. In our experiments, the performance of the proposed data augmentation is not degraded by the accuracy of the load prediction. For example, Fig. 3 presents very different outcomes of load prediction in two cases, i.e., Case 1 and Case 6, which does not affect the data augmentation significantly (see results in the following sections). To what extent the quality of the prediction impacts the detection accuracy will be left for future work.



(a) Bus 5 prediction in Case 1. (b) Bus 5 prediction in Case 6.

Fig. 3: Examples of load prediction based on LSTM model.

Fault Injection Component. For each operating condition, we implemented three-phase-to-ground faults by changing the resistance between a bus and a ground to 0.5 Ω. The faults of the two power systems were implemented with different simulation environments. We simulate the faults of the IEEE 9-bus system in a high-fidelity hardware-in-the-loop simulator from Opal RT; a total of 25.14 GB training data is generated. Avoiding the exploded size of training data in the IEEE 39-bus system (which can also reduce training overhead significantly), we simulate the faults of the IEEE 39-bus system in PowerWorld, a commercial simulator used in real utilities. Evaluations from the two different simulation environments show that the results are general and the benefits of the data-augmentation techniques are not affected by the specific simulation environment.

Based on these implementations, we present the evaluation results from these two cases in the following subsections.

C. Case Study 1: IEEE 9-Bus

Implementation. The IEEE 9-Bus implementation used OpalRT and Hypersim to simulate normal and fault conditions. For each operational condition, we have simulated 20 faults per bus with various duration. Consequently, each evaluation case includes a total of 16,000 conditions (including benign cases without faults) as historical data, and 4,000 conditions as the testing data set. In the OpalRT simulations, the ground fault was randomly triggered between 0.03 and 0.06 seconds and cleared after 0.005-0.035 seconds.

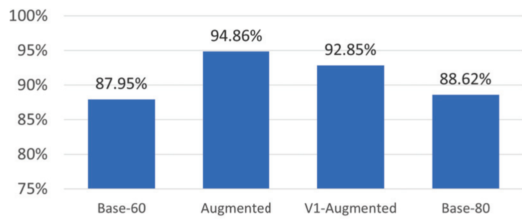


Fig. 4: Comparisons of fault detection accuracy under different training scenarios for the 9 Bus model.

Results. In Fig. 4, we compare the fault detection accuracy of various training scenarios, averaged across all six evaluation cases. We quantify fault detection accuracy by comparing the number of faults detected by the CNN model and the actual number of injected faults. By comparing the results in Base-60 and Base-80 scenarios, we observe that simply adding more historical conditions in the training data set can introduce few benefits in the fault detection. The main reason is that operating conditions in a 80-hour period experiences small changes. On the contrary, both Augmented and V1-Augmented scenarios can significantly increase the performance of the fault detection by 6.24% and 4.23% respectively. Note that the overhead of data collection is a critical factor to train deep learning models; this result can be significantly beneficial by increasing model accuracy without any additional efforts to collect more historical data.

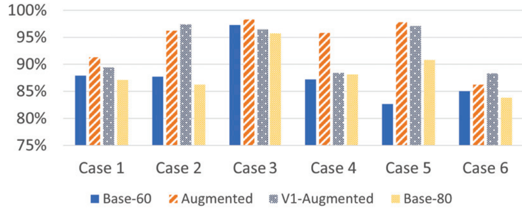


Fig. 5: Comparisons of fault detection accuracy in different evaluation cases in the 9 Bus model.

In Fig. 5, we further breakdown the results in different evaluation cases and observe two exciting and important findings. First, the data-augmentation can increase the performance of fault detection relying exclusively on historical data, regardless of its original performance. This observation is confirmed by both Augmented and V1-Augmented scenarios. Using augmented training data sets can increase the accuracy up to 12%. Furthermore, even though the size of the training data set used in the V1-augmented scenario is smaller than the size of the training data set used in the Augmented scenario, the detection accuracy still shows a dramatic improvement compared to the historical only training experiment in the Base-60 and Base-80 scenarios.

The second finding is that data-augmentation improves performance even in the case where the prediction accuracy is lower. In Case 6, the predicted load demands significantly deviate from the actual load demands, compared to the other five test cases (see Fig. 3). Even under this situation, fault detection accuracy improves by 4% to 8%, suggesting the data augmentation algorithm is not sensitive to prediction accuracy, likely because it enriches training data even though these new operating states do not exactly match the actual future state.

D. Case Study 2: IEEE 39 Bus Model

Implementation. The IEEE 39-Bus model was implemented in PowerWorld simulator. Specifically, the faults were simulated in each bus for the six evaluation cases. The duration of the faults varies from 0.01 second to 0.05 second with the step size of 0.01 second. Consequently, we collected a total of 16,000 historical data points and 4,000 predicted data points.

Results. In Fig. 6, we can observe similar improvements of fault detection in this larger system due to the data-augmentation. Including the prediction in training improves detection accuracy by 9.8% on average over the Base-60 scenario and 0.9% over the base-80 scenario. Note that the detection based on the Base-80 scenario already achieved 95.8% accuracy, probably due to the reduction in variability of fault simulation in the PowerWorld simulation. With 25% less data used in training, fault detection based on the V1-Augmented scenario can still achieve comparable accuracy (around 94%) to the detection accuracy based on the Base-80 scenario.

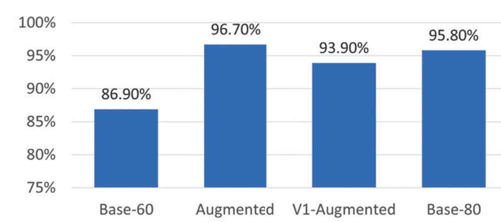


Fig. 6: Comparisons of average fault detection accuracy under different training scenarios for the 39 Bus model.

We further breakdown the evaluation results in six evaluation cases in Fig. 7. The CNN model used in fault detection behaves differently for IEEE 39-bus system; increasing the size of the training data set can increase the model accuracy. However, comparing the results between the V1-Augmented and Base-60 scenarios (whose training data set has the same size), we can still increase the detection accuracy. Specifically, we achieve 2.3% improvement in Case 5 (which observe accurate detection in Base-60 scenario) and 18.5% in Case 1. In other words, the proposed data augmentation methods can also be helpful when increasing the size of training data set becomes challenging.

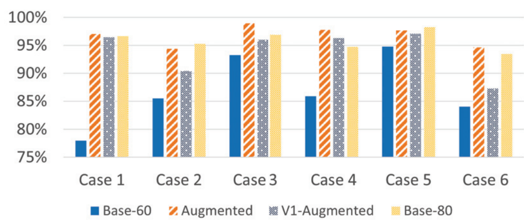


Fig. 7: Comparisons of fault detection accuracy under different evaluation cases in the 39 Bus model.

V. Conclusion

In this paper, we present a prediction-based data augmentation method that can increase fault detection accuracy in power grids. To "learn from the future," we create a training set integrating the knowledge of both future and faulty operating conditions with minor training overhead. Our implementation integrates the load prediction built on LSTM and two simulation environments, OPAL-RT and PowerWorld. Based on evaluations of four training scenarios and six evaluation cases in two IEEE test systems, utilizing the augmented training data set can increase the accuracy of data-driven fault detection by 8.6% on average, compared to the results trained based on historical data only.

With the promising results, we will focus on two aspects in future work: (i) systematic analysis on the impact of the amount of future data on fault detection and other data-driven applications and (ii) scenarios where the performance of load-prediction is downgraded, which may be caused by load anomalies or data compromise caused by malicious actors.

VI. Acknowledgement

This work was supported by the Office of Naval Research under Contract no. N00014-20-C-1096. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the view of the Office of Naval Research.

This material is based upon work partially supported by the National Science Foundation under Award No. CNS-2144513. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] "Puerto Rico faces blackout after cyberattack, fire" Aljazeera [Online]. Available: https://www.aljazeera.com/news/2021/6/11/electric-company-reports-fire-cyber-attack-in-puerto-rico.

[2] V. Ford, A. Siraj and W. Eberle, "Smart grid energy fraud detection using artificial neural networks," *2014 IEEE Symposium on Computational Intelligence Applications in Smart Grid* (CIASG), 2014, pp. 1-6.

[3] Xinlin Wang, Sung-Hoon Ahn, "Real-time prediction and anomaly detection of electrical load in a residential community," *Applied Energy*, vol. 259, 2020, pp. 114145.

[4] R. Moghaddass and J. Wang, "A hierarchical framework for smart grid anomaly detection using large-scale smart meter data," in *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 5820-5830, Nov. 2018.

[5] H. Jiang, J. J. Zhang, W. Gao and Z. Wu, "Fault detection, identification, and location in smart grid based on data-driven computational methods," *IEEE Trans. Smart Grid*, vol. 5, no. 6, Nov. 2014.

[6] Y. Liu, A. S. Meliopoulos, Z. Tan, L. Sun, and R. Fan, "Dynamic state estimation-based fault locating on transmission lines," *IET Generation, Transmission & Distribution*, vol. 11, no. 17, pp. 4184–4192, 2017.

[7] A R. Langeroudi, "Preventative high impedance fault detection using distribution system state estimation," *Electric Power Systems Research*, vol. 186, 2020.

[8] E. B. M. Tayeb and O. A. A. A. Rhim, "Transmission line faults detection, classification and location using artificial neural network," in *Proc. of 2011 International Conference & Utility Exhibition on Power and Energy Systems: Issues and Prospects for Asia (ICUE)*, pp. 1-5.

[9] M. He and J. Zhang, "A dependency graph approach for fault detection and localization towards secure smart grid," *IEEE Trans. Smart Grid*, vol. 2, no. 2, pp. 342-351, June 2011.

[10] E. Hossain, I. Khan, F. Un-Noor, S. S. Sikander and M. S. H. Sunny, "Application of big data and machine learning in smart grid, and associated security concerns: a review," *IEEE Access*, vol. 7, 2019.

[11] S. Zhang, Y. Wang, M. Liu and Z. Bao, "Data-based line trip fault prediction in power systems using LSTM networks and SVM," in *IEEE Access*, vol. 6, pp. 7675-7686, 2018.

[12] S. Iqbal et al., "Brain tumor segmentation in multi-spectral MRI using convolutional neural networks (CNN)," *Microscopy Research and Technique*, vol. 81, no. 4, 2018, pp. 419–427.

[13] H. M. Rouzbahani, H. Karimipour and L. Lei, "An ensemble deep convolutional neural network model for electricity theft detection in smart grids," in *Proc. of 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 3637-3642.

[14] X. Niu, J. Li, J. Sun and K. Tomsovic, "Dynamic detection of false data injection attack in smart grid using deep learning," in *Proc. of 2019 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pp. 1-6.

[15] H. Hassani, M. Farajzadeh-Zanjani, R. Razavi-Far, M. Saif and V. Palade, "Design of a cost-effective deep convolutional neural network–based scheme for diagnosing faults in smart grids," in *Proc. of 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, 2019, pp. 1420-1425.

[16] W. Danilczy, Y. Sun, and H. He, "Smart grid anomaly detection using a deep learning digital twin," in *Proc. of 52nd North American Power Symposium (NAPS)*, 2021.

[17] M. Al Olaimat, D. Lee, Y. Kim, J. Kim, and J. Kim, "A learning-based data augmentation for network anomaly detection," in *Proc. of 29th International Conference on Computer Communications and Networks (ICCCN)*, 2020, pp. 1-10.

[18] Z. A. Nazi, A. Biswas, M. A. Rayhan and T. Azad Abir, "Classification of ECG signals by dot residual LSTM network with data augmentation for anomaly detection," in *Proc. of 22nd International Conference on Computer and Information Technology (ICCIT)*, 2019, pp. 1-5.

[19] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu and Y. Zhang, "Short-Term residential load forecasting based on LSTM recurrent neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 841-851, Jan. 2019.

[20] L. Hernandez, et al., "A survey on electric power demand forecasting: future trends in smart grids, microgrids and smart buildings," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, third quarter, 2014.

[21] Y. Hua, Z. Zhao, R. Li, X. Chen, Z. Liu and H. Zhang, "Deep learning with long short-term memory for time series prediction," *IEEE Communications Magazine*, vol. 57, no. 6, pp. 114-119, June 2019.

[22] X. Wang, T. Zhao, H. Liu and R. He, "Power consumption predicting and anomaly detection based on long short-term memory neural network," in *Proc. of 4th IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, 2019, pp. 487-491.

[23] W. Danilczyk, Y. Sun, and H. He "Angel: an intelligent digital twin framework for microgrid security," *in Proc. of 2019 North American Power Symposium (NAPS)*, 2019, pp. 1-6.

[24] A. B. Birchfield, T. Xu, K.M. Gegner, K.S. Shetye, T.J. Overbye, "Grid structural characteristics as validation criteria for synthetic networks," *IEEE Trans. Power Systems*, vol. 32, no. 4, pp. 3258-3265, July 2017.