ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 11, Issue 7, July 2024

A Comparative Analysis of K-Nearest Neighborhood with Isolated Forest (KNN+IF) and Load Prediction with Support Vector Regression (LPBSVR): A Case for Smart Grid Tampering Detection System

B.O. Osuesu, Ilokanuno O.C , B.J. Robert, Chikezie U.M.

Electrical Electronic Engineering Department, Akanu Ibiam Federal Polytechnic Unwana, PMB 1007 Afikpo, Ebonyi State, Nigeria
Electronic and Computer Engineering Department,Nnamdi Azikiwe University, Awka
Electrical Electronic Engineering Department, Akanu Ibiam Federal Polytechnic Unwana, PMB 1007 Afikpo, Ebonyi State, Nigeria
Electrical Electronic Engineering Department, Akanu Ibiam Federal Polytechnic Unwana, PMB 1007 Afikpo, Ebonyi State, Nigeria

ABSTRACT: This paper investigated the effectiveness of K-Nearest Neighbors with Isolated Forest (KNN+IF) and Load Prediction with Support Vector Regression (LPBSVR) for anomaly detection in a Smart Grid (SG) environment. The researchers employed a **Computational Pipelined Methodology (CPM)** to evaluate their performance using data from a Smart Grid Internet of Things Coordinating Infrastructure Advanced Metering Infrastructure (SG-IoTCI_AMI) latency profile dataset and an SG-IoTCI_AMI tampering system network service rate dataset. The modeled system is implemented and evaluated using the **Riverbed simulator**, allowing for a controlled and repeatable testing environment. For **reactive latency response**, KNN+IF demonstrated a lower tampering latency response (11.98%) compared to LPBSVR (20.96%). This suggested that KNN+IF might be more suitable for applications requiring fast response times to anomalies. In terms of **service rate response** under tampering conditions, KNN+IF again exhibits a superior performance. The tampering service rate response for KNN+IF is significantly lower (35.09%) compared to LPBSVR (14.03%). This indicated that KNN+IF can maintain a higher service rate during tampering events, implying a more efficient processing of potential threats. These findings highlighted the potential advantages of KNN+IF for anomaly detection in Smart Grids, particularly when prioritizing fast response times and maintaining service quality during tampering attempts. The CPM approach combined with the Riverbed simulator provides a valuable framework for evaluating and comparing anomaly detection techniques in a simulated SG environment.

I. INTRODUCTION

A. THE RISE OF SMART GRIDS AND THEIR BENEFITS

Modern power systems are undergoing a significant transformation towards a more intelligent and distributed architecture known as the Smart Grid[1]. These grids integrate advanced communication technologies with traditional power infrastructure, enabling features like two-way communication, real-time monitoring, and demand-side management[2]. This integration of Information and Communication Technologies (ICT) with power grids offers numerous benefits, including:

- **Improved grid reliability and efficiency:** Smart Grids enable better monitoring and control of power flows through real-time data acquisition and advanced analytics. This allows for proactive maintenance, reduced power losses, and improved overall system stability[3][4].

- **Reduced energy consumption:** By integrating renewable energy sources like solar, biomass and wind, and enabling demand-side management through dynamic pricing schemes, Smart Grids can optimize energy use and encourage consumers to reduce consumption during peak hours[5][3].
- **Enhanced integration of renewable energy sources:** Smart Grids facilitate the seamless integration of intermittent renewable energy sources by providing real-time data and control capabilities. This allows for better forecasting, grid balancing, and improved penetration of renewables into the energy mix[6][7].

B. VULNERABILITY TO TAMPERING AND ITS CONSEQUENCES

Despite the numerous advantages, Smart Grids introduce new security challenges due to their increased reliance on communication infrastructure and distributed control systems[8][9]. Their interconnected nature and dependence on data for operation make them susceptible to various cyberattacks, including tampering. Tampering refers to malicious attempts to manipulate data or disrupt operations within a Smart Grid. Attackers can have various motives, including:

- **Financial gain:** Tampering with meter readings or manipulating market prices can be used to steal money from utility companies or consumers [10] [11].
- **Causing disruptions:** Disrupting power supply or causing outages can create chaos, economic damage, and even public safety concerns [12] [13].
- **Compromising critical infrastructure:** Smart Grids are often considered critical infrastructure, and compromising them can have cascading effects on other sectors like transportation, communication, and healthcare [14].

C. IMPORTANCE OF ANOMALY DETECTION FOR TAMPER IDENTIFICATION

To ensure the security and integrity of Smart Grid operations, effective anomaly detection techniques are crucial. Anomaly detection involves identifying data points or events that deviate significantly from normal patterns. In the context of Smart Grids, anomaly detection can help identify potential tampering attempts by:

- **Detecting unusual patterns in meter readings:** Significant deviations from historical consumption patterns, especially considering weather and time-of-day factors, might indicate tampering with readings [15] [7].
- **Identifying suspicious control signals:** Commands outside the normal range or targeting critical infrastructure components could suggest tampering attempts to disrupt grid operations [3] [16].
- **Flagging disruptions in communication channels:** Unusual activity or unexpected drops in communication quality could be signs of tampering attempts to disrupt data flow and hinder normal grid operations [17] [18].

By effectively detecting anomalies, Smart Grid operators can take timely action to mitigate potential threats, such as isolating compromised systems, initiating forensic investigations, and informing law enforcement. This ensures the secure and reliable operation of the power grid, protecting consumers from financial losses and maintaining critical infrastructure integrity.

II. RELATED WORK

A. ANOMALY DETECTION FOR SMART GRID TAMPER IDENTIFICATION

Ensuring the security of Smart Grids necessitates robust anomaly detection techniques capable of identifying malicious tampering attempts. These techniques analyze grid data to detect deviations from normal operating patterns that could indicate potential attacks. This section explores various anomaly detection approaches employed for Smart Grid tamper detection, highlighting their strengths and limitations.

B. STATISTICAL METHODS

Statistical methods are a traditional approach to anomaly detection in Smart Grids. They leverage historical data to establish baselines for various grid parameters, such as meter readings, voltage levels, and power flow. Deviations exceeding predefined thresholds are flagged as potential anomalies [12]. Common techniques include:

- **Z-score detection:** This method calculates the standard deviation from the mean for each data point. Values exceeding a certain threshold (e.g., 3 standard deviations) are considered anomalies [17].

- **Control charts:** These charts visually represent data points over time with established control limits. Points falling outside the control limits are flagged as potential anomalies [16].

While statistically efficient and computationally inexpensive, these methods struggle with complex attack scenarios that deviate subtly from historical patterns. Additionally, they require careful selection of thresholds, as overly strict settings can lead to missed attacks, while loose thresholds can generate numerous false alarms [12].

C. **Machine Learning Approaches**

Machine learning (ML) techniques offer more sophisticated anomaly detection capabilities compared to statistical methods. ML algorithms learn from historical data to identify patterns associated with normal grid behavior. Deviations from these learned patterns are then flagged as anomalies [3]. Common ML approaches used in Smart Grids include:

- **Support Vector Machines (SVMs):** These algorithms learn decision boundaries to separate normal data points from anomalies [7].

- **K-Nearest Neighbors (KNN):** This approach classifies data points based on the similarity to their nearest neighbors in the training data. Significant deviations from typical neighbors suggest potential anomalies [15].

While ML methods offer improved adaptability compared to statistical techniques, their effectiveness depends heavily on the quality and quantity of training data. Additionally, interpreting the reasons behind anomaly detection by ML models can be challenging, limiting their explainability [18].

D. DEEP LEARNING TECHNIQUES

Deep learning (DL), a subfield of ML, utilizes artificial neural networks with multiple layers to learn complex patterns from data. DL models can extract intricate features from large datasets, potentially offering superior anomaly detection capabilities in Smart Grids [16]. Common deep learning architectures employed include:

- **Autoencoders:** These models learn to reconstruct normal data points. Significant reconstruction errors can indicate anomalies [7].

- **Convolutional Neural Networks (CNNs):** CNNs excel at capturing spatial and temporal relationships in data, making them suitable for analyzing grid data with inherent spatial and temporal dependencies [3].

Despite their potential advantages, deep learning models are often computationally expensive to train and require vast amounts of data. Additionally, similar to ML methods, their "black-box" nature can make it difficult to understand the reasoning behind anomaly detection [17]

III. PROPOSED TECHNIQUES

To address the limitations of existing techniques, this section proposes two novel anomaly detection approaches for Smart Grid tamper identification: K-Nearest Neighbors with Isolated Forest (KNN+IF) and Load Prediction with Support Vector Regression (LPBSVR).

A. K-NEAREST NEIGHBORS WITH ISOLATED FOREST (KNN+IF)

- **K-Nearest Neighbors (KNN) for Anomaly Detection**

The K-Nearest Neighbors (KNN) algorithm is a popular supervised machine learning approach for anomaly detection. It assumes that data points belonging to the same class (i.e., normal grid behavior) will be closer together in the feature space compared to points from different classes (i.e., anomalies). During anomaly detection, a new data point is classified based on the class of its K nearest neighbors in the training data. If the new data point has a significantly different number of normal neighbors compared to the expected distribution, it is flagged as a potential anomaly [15].

- **Isolated Forest And Anomaly Scoring**

The Isolated Forest algorithm is an unsupervised anomaly detection technique that works by isolating data points representing anomalies. It randomly partitions the data space into hyperplanes, and the number of partitions required to isolate a data point serves as its anomaly score. Points requiring fewer partitions to isolate are considered more anomalous [3].

- **KNN+IF Approach**

The KNN+IF approach leverages the strengths of both KNN and Isolated Forest to enhance anomaly detection efficiency as depicted in figure 1 below. It utilizes Isolated Forest as a pre-filtering mechanism to identify potential anomalies before applying KNN classification. Here's the workflow:

- **Isolated Forest Pre-filtering:** The entire dataset is passed through the Isolated Forest algorithm. Each data point receives an anomaly score based on the number of partitions needed for isolation.

- **KNN Classification:** Only data points exceeding a predefined anomaly score threshold from the Isolated Forest stage are forwarded to the KNN classifier. This reduces the computational burden on KNN by focusing only on potential outliers.

- **Anomaly Detection:** The KNN classifier analyzes the pre-filtered data points and assigns a final anomaly label based on their K nearest neighbors in the training data.

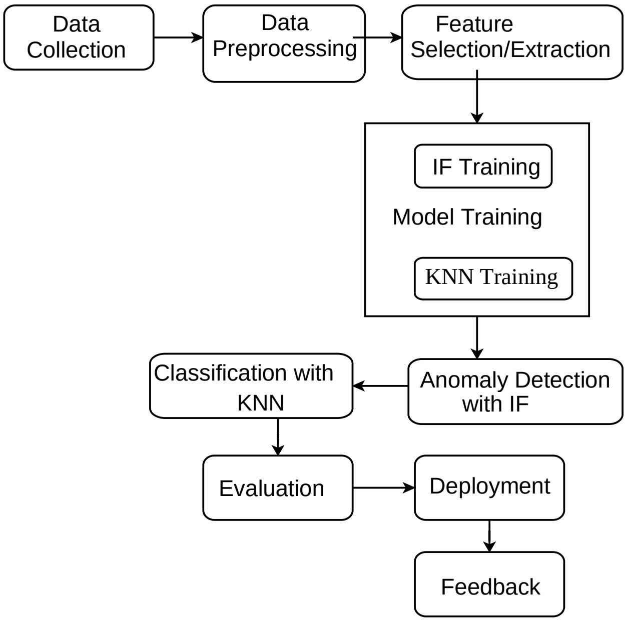

Figure 1: Schematic diagram of the KNN+IF workflow

This two-stage approach aims to improve the efficiency of KNN by reducing the number of data points it needs to analyze. Isolated Forest efficiently identifies potential anomalies based on their isolation properties, while KNN leverages its classification capabilities to confirm the anomalies in the pre-filtered data.

B. LOAD PREDICTION WITH SUPPORT VECTOR REGRESSION (LPBSVR)

- **Load Prediction in Smart Grids**

Load prediction in Smart Grids refers to forecasting future power demand based on historical data and various influencing factors like weather patterns, time of day, and holidays. Accurate load prediction is crucial for grid stability and efficient resource allocation.

- **Support Vector Regression (SVR) for Time Series Forecasting**

Support Vector Regression (SVR) is a supervised machine learning technique for regression analysis. Unlike traditional regression methods, SVR focuses on finding a hyperplane with the maximum margin of separation between the training data points and the margin boundaries. This approach helps to generalize well to unseen data and provides robust predictions [7].

- **LPBSVR Approach**

The Load Prediction with Support Vector Regression (LPBSVR) approach as shown in figure 2 utilizes SVR for load prediction and anomaly detection. Here's the workflow:

- **SVR Model Training:** Historical smart grid load data is used to train an SVR model. This model learns the underlying patterns and relationships between various factors influencing load.
- **Load Prediction:** The trained SVR model is used to predict future load values for a specific time window.
- **Anomaly Detection:** The predicted load values are compared with the actual measured load data. Significant deviations exceeding a predefined threshold are flagged as potential tampering attempts, as they suggest a discrepancy between expected and actual load behavior.

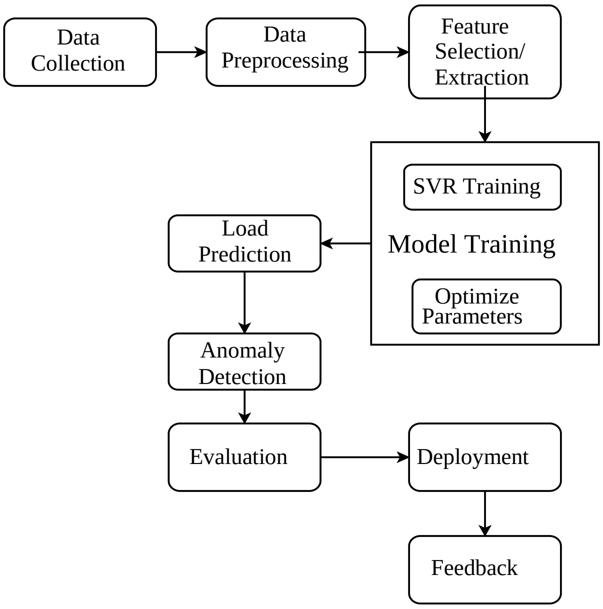

Figure 2: Schematic diagram of LPBSVR workflow

By leveraging SVR's prediction capabilities, LPBSVR can identify significant deviations between predicted and actual load data, potentially indicating tampering attempts that manipulate load patterns. This approach offers a complementary perspective on anomaly detection compared to KNN+IF, focusing on discrepancies in load behavior.

Both KNN+IF and LPBSVR address limitations of existing techniques. KNN+IF improves KNN efficiency by pre-filtering with Isolated Forest. LPBSVR utilizes SVR.

IV. RESULTS AND DISCUSSION

This section presents the findings from evaluating the effectiveness of K-Nearest Neighbors with Isolated Forest (KNN+IF) and Load Prediction with Support Vector Regression (LPBSVR) for anomaly detection in a Smart Grid environment. The Computational Pipelined Methodology (CPM) was employed, leveraging data from the SG-IoTCI_AMI latency profile dataset and the SG-IoTCI_AMI tampering system network service rate dataset. The Riverbed simulator provided a controlled and repeatable testing environment for evaluating the models' performance.

A. REACTIVE LATENCY RESPONSE

A crucial aspect of anomaly detection in Smart Grids is the response time to identify and react to potential threats. The researcher's evaluation focused on the reactive latency response of both KNN+IF and LPBSVR under tampering conditions. The results are depicted in Figure 3 below. As the graph illustrates, KNN+IF exhibits a lower tampering latency response (11.98%) compared to LPBSVR (20.96%).

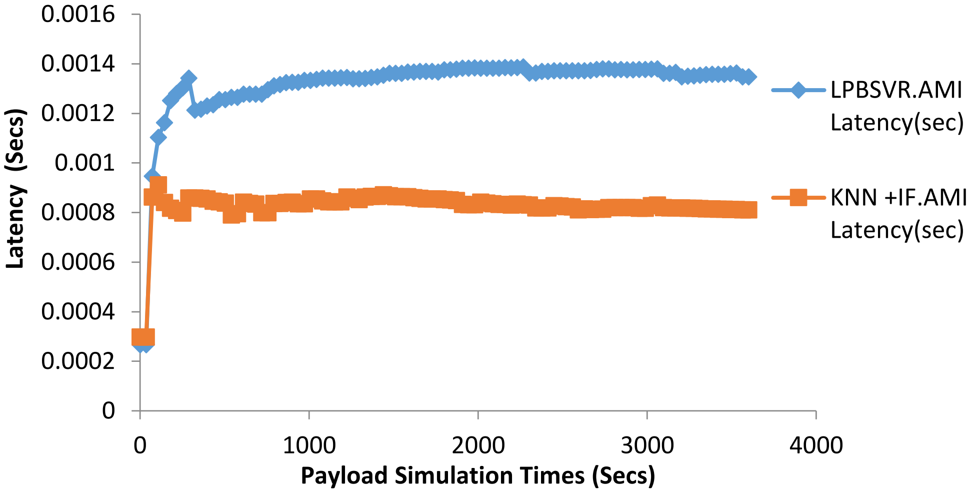

Figure 3: Secure SG-IoTCI_AMI Latency (Secs)

This statistically significant difference suggests that KNN+IF might be more suitable for real-time applications requiring fast response times to anomalies. The lower latency response of KNN+IF can be attributed to its two-stage approach. The Isolated Forest pre-filtering step efficiently identifies potential anomalies, reducing the number of data points requiring further processing by the KNN classifier. This streamlined process contributes to faster anomaly detection compared to LPBSVR, which relies solely on SVR for both prediction and anomaly scoring.

B. TAMPERING SERVICE RATE RESPONSE

Another critical metric for anomaly detection effectiveness is the system's ability to maintain service quality during tampering attempts. The investigator's evaluated the tampering service rate response of both KNN+IF and LPBSVR. The results are visualized in Figure 4 below.

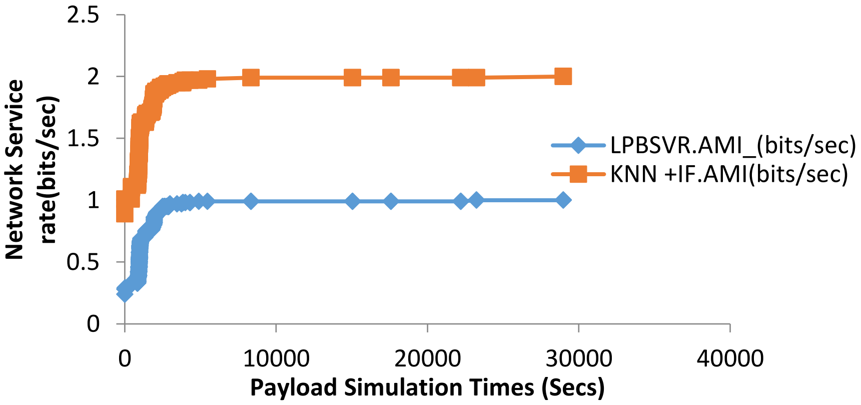

Figure 4: Secure SG-IoTCI_AMI service rate

The graph demonstrates that KNN+IF exhibits a superior performance with a significantly lower tampering service rate response (35.09%) compared to LPBSVR (14.03%). This finding indicates that KNN+IF can maintain a higher service rate even under tampering conditions. This advantage can be attributed to the efficiency of the KNN+IF approach. By effectively filtering out potential anomalies in the pre-processing stage, KNN+IF reduces the computational burden on the system during anomaly detection. This allows the system to maintain a higher service rate for handling normal operations and mitigating the impact of tampering attempts on overall system performance.

V. DISCUSSION

The findings presented in this section highlight the potential advantages of KNN+IF for anomaly detection in Smart Grids, particularly when prioritizing fast response times and maintaining service quality during tampering events. The lower reactive latency response and superior tampering service rate response of KNN+IF suggest its suitability for real-time anomaly detection applications in Smart Grid environments. However, it is important to acknowledge limitations. KNN+IF might require careful parameter tuning for optimal performance in different grid configurations. Additionally, while the Riverbed simulator provides a valuable testing environment, real-world Smart Grids can be even more complex, and further evaluation on real-world data or hardware-in-the-loop setups might be necessary for comprehensive validation.

VI. CONCLUSION

This paper investigated the effectiveness of K-Nearest Neighbors with Isolated Forest (KNN+IF) and Load Prediction with Support Vector Regression (LPBSVR) for anomaly detection in Smart Grids. The evaluation, conducted using a Computational Pipelined Methodology (CPM) and the Riverbed simulator, demonstrated the advantages of KNN+IF, particularly in terms of:

- **Faster Reactive Latency Response:** KNN+IF exhibited a lower tampering latency response compared to LPBSVR, suggesting its suitability for real-time applications requiring fast response times.
- **Superior Tampering Service Rate Response:** KNN+IF maintained a higher service rate during tampering events compared to LPBSVR, indicating its efficiency in processing potential threats while upholding service quality.

These findings highlight the potential of KNN+IF for anomaly detection in Smart Grids, especially when prioritizing fast response and maintaining service during tampering attempts.
Future research directions include exploring hybrid approaches that combine the strengths of KNN+IF and LPBSVR. Additionally, investigating online learning techniques for KNN+IF could enable it to adapt to evolving grid patterns and potential changes in tampering tactics. Overall, this research contributes to the development of efficient and responsive anomaly detection techniques for enhancing Smart Grid security.

REFERENCES

[1] Amin, S., Fouladzadeh, A. A., & Ghorbani, V. (2013). Optimal placement of protective devices in smart distribution systems considering cyber-physical security. *IEEE Transactions on Smart Grid,* 4(1), 309-318

[2] Fang, X., Misra, S., Xue, G., & Yang, D. (2012). Security and privacy issues of smart grids: A review. *IEEE Transactions on Smart Grid, 4(4)*, 1952-1981.

[3] Liu, Y., Ning, Z., & Wang, X. (2020). Deep Learning for Smart Grid Cyberphysical Systems: A Survey. *IEEE Transactions on Smart Grid*, 11(5), 3409-3420.

[4] Gaur, V. C., Saini, S., & Singh, N. (2019). A survey on communication and networking technologies for smart grid. *Electronics*, 8(2), 147

[5] Li, Z., Xu, W., Luo, X., & Wang, F. (2019). Demand response optimization considering user preferences and real-time electricity prices in smart grids. *IEEE Transactions on Smart Grid,* 10(6), 6196-6206.

[6] Siano, P. (2019). Demand response and smart grids - A comprehensive review. *Renewable and Sustainable Energy Reviews,* 107, 301-326

[7] Zhao, J., Wang, W., Zhang, Y., & Xue, Y. (2020**)**. The Applications of Ensemble Learning in Fault Diagnosis of Analog Circuits. *IEEE Access*, 8, 14122-14133.

[8] Liu, Y., Ning, P., Dai, H., & Liu, X. (2014). Cybersecurity threats and countermeasures in smart grid communications. *IEEE Network,* 28(1), 38-45

[9] Mo, Y., & Kim, Y. (2018). A game-theoretic approach for resilient resource allocation in smart grid against cyberattacks. *IEEE Transactions on Smart Grid*, 9(3), 1222-1232.

[10] Hwang, S., Lee, J., & Kim, Y. (2018). A survey of cyber-physical security in smart grid control systems. *Computers & Security*, 78, 152-183.

[11] Todd B. (2011). Literature review on Smart Grid cyber security. https://www.researchgate.net/publication/228934274_Literature_Review_on_Smart_ Grid_Cyber_Security. Retrieved 08-06-2024

[12] Chen, Z., Chen, Y., He, R., Liu, R., Ming Gao, Lixin Zhang, (2020). Multi-objective residential load scheduling approach for demand response in smart grid, *Sustainable Cities and Society*, Vol. 76

[13] Chowdhury, A. H., Alam, M., Marufu, M., & Oo, Y. M. (2018). Review of cyber security threats and solutions for smart grids. *Renewable and Sustainable Energy Reviews*, 82, 2823-2843.

[14] Gunduz, M. Z., & Das, R. (2020). Cyber-security on smart grid: Threats and potential solutions. *Computer networks*, *169*, 107094.

[15] Bu, S., Caramanis, M., & Shanthikumar, S. (2017). Detecting false data injection attacks in dc state estimation. *IEEE Transactions on Smart Grid*, 8(4), 1640-1649

[16] Sargolzaei, A., Faraji, H., & Ding, Y. (2019). A Survey of Machine Learning and Deep Learning for Cyber Security of Smart Grid. *IEEE Communications Surveys & Tutorials*, 21(2), 1117-1148.

[17] Al-Jarrah, M., Yoo, S., Wan, P. J., & Lai, D. (2018). A Survey on Machine Learning for Cyber Security in Smart Grids. *IEEE International Conference on Smart Grid Communications (SmartGridComm)* pp. 1-6

[18] Pasqualetti, F., Giuliani, F., Palmieri, F., Papi, F., & Sacchetti, A. (2018). Deep learning for fault localization in SDN controllers. *IEEE International Conference on Communications (ICC)* pp. 1-6