ISCADA: Towards a Framework for Interpretable Fault Prediction in Smart Electrical Grids

Carmelo Ardito, Yashar Deldjoo, Eugenio Di Sciascio, Fatemeh Nazary, Gianluca Sapienza

ISCADA: Towards a Framework for Interpretable Fault Prediction in Smart Electrical Grids

Carmelo Ardito[1][0000-0001-8993-9855], Yashar Deldjoo[1][0000-0002-6767-358X], Eugenio Di Sciascio[1][0000-0002-5484-9945], Fatemeh Nazary[1][0000-0002-6683-9453]⋆, and Gianluca Sapienza[2]

[1] Politecnico di Bari, Italy
firstname.lastname@poliba.it
[2] SmartGridLab e-distribuzione S.p.A., via Rubattino 84 - Milano, Italy
gianluca.sapienza@e-distribuzione.com

Abstract. This paper reports ongoing research for the definition of a data-driven self-healing system using machine learning (ML) techniques that can perform automatic and timely detection of fault types and locations. Specifically, the proposed method makes use of spectrogram-based CNN modeling of the 3-phase voltage signals. Furthermore, to keep human operators informed about why certain decisions were made, i.e., to facilitate the interpretability of the black-box ML model, we propose a novel explanation approach that highlight regions in the input spectrogram that contributed the most for the prediction task at hand (e.g., fault type or location) - or *visual explanation*.

Keywords: Self-healing system · Interpretability · Fault prediction.

1 Introduction and Context

Electrical grids are susceptible to a variety of electrical abnormalities, failures, and security threats, which if not tackled abruptly, in some cases they can leave a devastating impact on lives and the country's critical industries, leading to a national dilemma as the result of a cascading effect [5, 6]. Over the last decade, the problem of automatic fault detection has been studied from various viewing angles using a combination of tools and techniques from computer science, electrical engineering, statistics, and artificial intelligence (AI) and in particular using automatic data-driven machine learning (ML) algorithms [2, 7]. From a software point of view, the use of SCADA (Supervisory Control And Data Acquisition) has dominated the electrical industry. SCADA is a control system architecture composed of computers, networked data communications, and graphical user interfaces (GUIs) for high-level process supervisory management tasks [4]. The users of the SCADA, however, can be diverse ranging from electrical engineers to expert data-scientists, and lower professional-level workers.

⋆ Authors are listed in alphabetical order. Corresponding author: Fatemeh Nazary

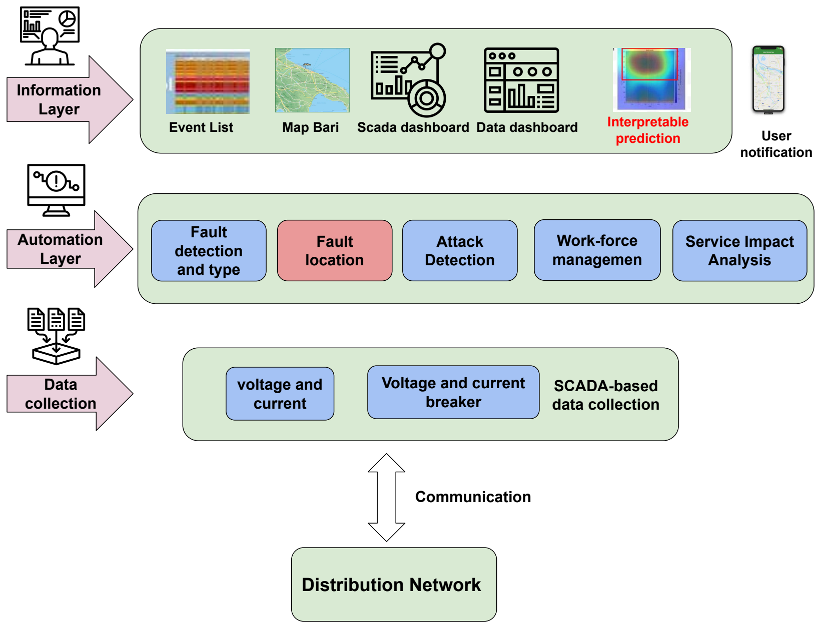

Fig. 1. The 3-layered management architecture of an electricity provider network. The red modules show our contribution.

2 Overview of the self-healing management system of an electricity provider workflow

We model the self-healing management system of an electricity distributor as shown in Figure 1 according to hierarchical architecture, where the lowest layer is the closest level to the input data (i.e., raw electrical signals), and the highest level is semantic data that are supposed to be understandable to the human observer, thus **bottom (data)→top (user)**. We describe the functionality of each layer below:

1. **Data-collection level:** The main role of the data collection layer is the acquisition of electrical, and logical data e.g., whether the fault breaker status is close (0) or open (1). The data collection task by an electricity provider is mainly managed by the SCADA (short for Supervisory control and data acquisition) system, which is a standard approach to different types of input data.
2. **Automation levels:** This level is responsible for automating the decision-making by the system. Fault classification is the functionality that is inside the operative center, however, traditionally it relies on the knowledge of specialists that analyze fault to understand if it is necessary to inspect the feeder. Attack detection (remote control part of smart grid), is mainly managed by the telecommunication network. The workforce management unit

relates to the crew who directly goes to check the problem. Service impact analysis deals with regulatory and quality of service. We have added an accurate data-driven location prediction system (where the latter is shown with **red** color) for isolating faulty zones or their precise location. We believe this can be a crucial part of modernized self-healing SGs since it expedites the dispatch of the crew members to the targeted (faulty) areas.
3. **Information level:** This level is responsible for reporting the monitoring status and fault decision outcomes to the user/operator. Practically, this is implemented in electricity providers using SCADA dashboard and data dashboard that can summarize data or give some detailed information to the user, depending on the type level of analysis he is interested in.

In the next section, we elaborate our contribution to the existing management self-healing system of the electricity provider.

3 Our proposed contribution

We propose *ISCADA*, interpretable SCADA that aims to augment fault prediction results in SGs with information that facilitates interpreting them to both experts and non-experts operators or users. To this end, we have built a deep-learning-based fault prediction system[3] that can provide accurate and interpretable fault type and location with relatively high precision. ISCADA extends the traditional SCADA by adding other components in the specific layers, as shown with red colors in Figure 1. They include a fault location detection (automation layer) and interpretable fault prediction visualization (data layer). We now describe the main idea and the functioning mechanism behind the proposed work.

As shown in Figure 2, starting from the raw 3-phase voltage signal collected from a hypothetical faulty area (zone A), we transform the data into a spectrogram representation, which will serve as a visual time-frequency representation of the original signal [3]. This representation is not only an efficient representation of the signal as it contains both time and frequency representation, but also a suitable technique for visual inspection of the fault prediction results according to time and frequency changes in the signal. For the fault prediction system, we build a data-driven classification system utilizing a powerful **deep convolutional neural network** (CNN). Furthermore, an approach based on gradient-weighted class activation mapping (Grad-CAM) [8] technique is utilized to highlight important regions in the spectrogram that can explain fault type and location, or **visual explanation**. The result of the explanation is a heatmap visualizing where on the spectrogram images the network was paying attention to. These can provide insightful visual cues to the human operator about the characteristics of the faults and can keep them in the loop.

A hypothetical extension of these results would involve how to turn such interpretable outcomes [1, 2] into an explanation for non-expert users. In our

[3] This work is currently under review at the IEEE Transaction of smart grids (SG)

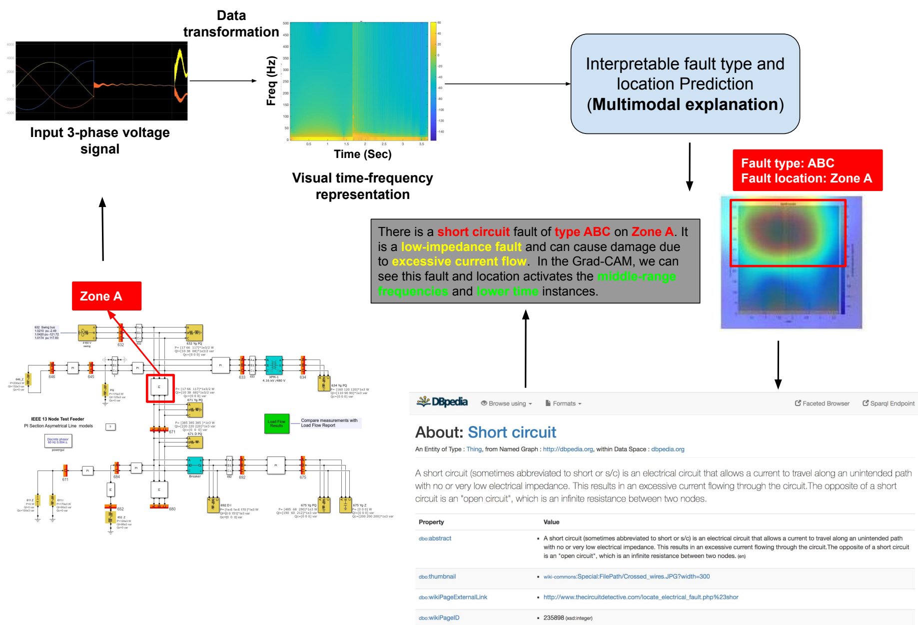

Fig. 2. Illustration of the principles behind the proposed intereptable fault prediction system

scenario, we can use the result of fault prediction as an input to a knowledge graph or other semantic repository of knowledge that can be also found freely on the web. An example of such freely available resource we can name of DBPedia, as shown in Figure 2. The results are presented to the user in terms of **multi-modal explanation** using natural language (results returned from the KG) and visual outcome (the result of the automated fault prediction system shown with the red zone over the spectrograms).

4 Conclusion

In this paper, we have reported the results of an ongoing research work aiming at the definition of an interpretable and accurate data-driven self-healing system using machine learning (ML). The end goal on the one hand is to perform automatic and timely detection of fault types and locations; on the other hand, we aim to make the prediction results interpretable to the user. We discussed briefly the main insights behind our proposed CNN-based modelling of fault spectrograms and the Grad-CAM technique, and how the latter can be used to make the decisions made by the system more transparent.

Acknowledgments

This work has been partially funded by *e-distribuzione S.p.A* company, Italy, through a PhD scholarship granted to Fatemeh Nazary.

References

1. Ardito, C., Deldjoo, Y., Di Sciascio, E., Nazary, F.: Interacting with features: Visual inspection of black-box fault type classification systems in electrical grids. In: Italian Workshop on Explainable Artificial Intelligence (XAI.it@AIxIA'20) (2020)
2. Ardito, C., Deldjoo, Y., Di Sciascio, E., Nazary, F.: Revisiting security threat on smart grids: Accurate and interpretable fault location prediction and type classification. In: The Italian Conference on CyberSecurity (ITASEC) (2021)
3. Birajdar, G.K., Patil, M.D.: Speech and music classification using spectrogram based statistical descriptors and extreme learning machine. Multim. Tools Appl. **78**(11), 15141–15168 (2019). https://doi.org/10.1007/s11042-018-6899-z, https://doi.org/10.1007/s11042-018-6899-z
4. Boyer, S.A.: SCADA: supervisory control and data acquisition. International Society of Automation (2009)
5. Burpee, D., Dabaghi, H., Jackson, L., Kwamena, F., Richter, J., Rusnov, T., Friedman, K., Mansueti, L., Meyer, D.: Us-canada power system outage task force: final report on the implementation of task force recommendations (2006)
6. Majidi, M., Etezadi-Amoli, M., Fadali, M.S.: A sparse-data-driven approach for fault location in transmission networks. IEEE Trans. Smart Grid **8**(2), 548–556 (2017). https://doi.org/10.1109/TSG.2015.2493545, https://doi.org/10.1109/TSG.2015.2493545
7. Raja, S., Fokoué, E.: Multi-stage fault warning for large electric grids using anomaly detection and machine learning. CoRR **abs/1903.06700** (2019), http://arxiv.org/abs/1903.06700
8. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision. pp. 618–626 (2017)