ML/AI Solutions for Protection and Control Applications and Expected Benefits

M. Kezunovic
Texas A&M University
m-kezunovic@tamu.edu
U.S.A.

Abstract: This paper explores the following questions: a) how variety of data about faults and outages may be utilized for automated analysis of such events, b) what will be the new decision-making framework for utility operators using the analysis results, and c) where would such solutions fit as the tools intended for protection and control (PAC) engineers? Those research questions and practical application aspects are explored through a couple of machine learning/artificial intelligence (ML/AI) use cases for transmission and distribution applications, namely: a) automated analysis of the faults using ML/AI on synchrophasor data aimed at transmission system operators that are responsible for real-time decision-making when the system experiences faults, and b) automated prediction of the outage state of risk (SoR) in distribution systems caused by environmental conditions around the feeders such as severe weather extremes, lightning strikes and vegetation growth aimed at asset and outage management engineers responsible for the system maintenance and restoration. Possible utilization of the results by PAC engineers are explored in each case.

Introduction

The power system monitoring, control and protection requirements are changing due to several emerging reasons: a) power systems dynamics are evolving with addition of inverter based resources (IBRs) at the transmission level with deployment of utility scale renewable generation such as solar and wind-based inverter-connected power plants, and distributed generation resources (DERs) at the distribution level with addition of third party owned renewable generation and both mobile (electrical vehicles) and stationary (battery) energy storage connected via inverters [1,2]; b) new field recording technology is also evolving by addition of synchrophasor systems and point-on-wave (POW) capturing systems across transmission and distribution [3,4], and c) the emergence of wealth of recorded data from legacy and new recording systems combined with wealth of environmental data is proliferating the use of ML/AI applications on such data to detect, classify and characterize the faults and predict outages [5,6].

The mentioned developments have raised an issue of how ML/AI may be used to enhance the decision-making capability of PAC engineers, and in what circumstances such tools may be used. In this context, we are expanding the analysis and decision making process to include: a) on-line real-time monitoring of the transmission system disturbances such as faults to aid transmission system operators in making timely decisions about the system control, and b) predictive capability of anticipating distribution system outages providing asset and outage management engineers and related repair and restoration crews with capability to prepare mitigation measures to reduce or even avoid outage impacts. Such developments go well beyond the traditional approach of automating the analysis of protective relaying and digital fault recording data [7] to complement the supervisory control and data acquisition (SCADA) recording since in our approach new data and new analysis methods are used [8,9].

In this paper, we focus on two recent trends in the uses of emerging data, and related ML/AI applications, namely: a) automated analysis of synchrophasor data, and b) outage state of risk (SoR) prediction using historical grid outage data and extensive environmental data. We are surveying recent developments by pointing to several recent studies where we have used utility historical data and other field data to create new results not utilized before [10-11]. The contribution of this paper is in defining new decision-making paradigm that PAC engineers may be facing in the future when performing their everyday tasks, and suggesting how their tasks may be expanded and enhanced using ML/AI-based solutions. We start with the background section where we explain how the legacy systems and related data may be expanded, then we show examples of ML/AI applications in the transmission and continue with a section on distribution application. The benefits and requirements of such solutions and list of references are given at the end.

Legacy solutions and new data analysis opportunity

Transmission applications

Traditional approaches to automated fault analysis in transmission systems focus on the use of data recorded in transmission substations. Figure 1 shows a typical transmission substation recording equipment used for fault analysis: fault locators (FL), digital fault recorders (DFRs), phasor measurement units (PMUs), digital protective relays (DPRs), remote terminal units (RTUs) of a supervisory control and data acquisition (SCADA) system and sequence of events recorders (SOE). They record samples of either analog (A) waveforms, (e.g., voltages and currents coming from the power system instrument transformers), or samples of status (S) data coming from contacts of the switching equipment, (e.g., "a" and "b" breaker contacts), or both A and S types. Some supporting communication and processing systems are developed to collect data from multiple substation devices or signals and store them in centralized locations. As an example, a centralized fault location (CFL) system may collect fault location data from multiple substations and keep it for the entire system at an engineering back office. Similarly, data from DFRs may be collected in substations from multiple DFR units and preprocessed in a local master station (LMS) and then communicated to corporate master stations (MS) for regional or company-wide collection and storage. Synchrophasor data is typically collected by PMUs in substations and forwarded in real time to phasor data concentrators (PDCs) that may be in a substation or at different company offices, and in many cases, passed over to the energy management system (EMS) for integration with SCADA data or direct display for control center operators. The RTU and SOE data are typically passed to the SCADA database to be used by legacy EMS functionalities and applications such as state estimation.

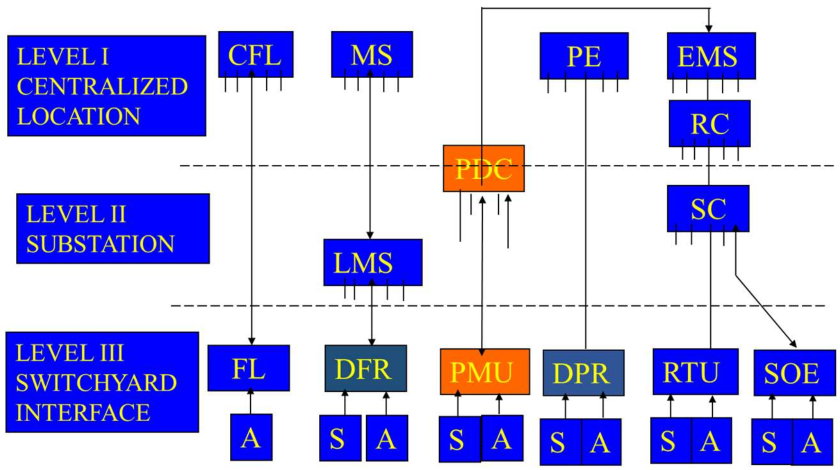


Figure 1 Substation recording equipment

Because such recording devices capture different data for different purposes, the users of such data are different groups in the utility company. An illustration of the recorded data properties that drive data uses is given in Table 1. Incidentally, an automated analysis can not only benefit the protection engineers but also the operators, particularly the Independent System Operators (ISOs), which only receive the PMU data streams but not the data captured by other substation devices. In reference to Figure 1, we are focusing on the automated analysis of PMU data streams using ML/AI-based data analytics as discussed next.

Table 1 Types and properties of substation recorded data

Source	Type of Data	Operating Mode	Reporting Frequency	Synchronization between devices
RTUs	Status RMS value	Report	Every several seconds	No
PMUs	Phasor	Report	Up to 60 times per second	Yes
ERDs	RMS /DFT /samples	Upon-request	Upon Request	No

RTUs: Remote Terminal Units;

PMUs: Phasor Measurement Units;

ERDs: Event Reporting Devices such as DFRs and DPRs

Distribution applications

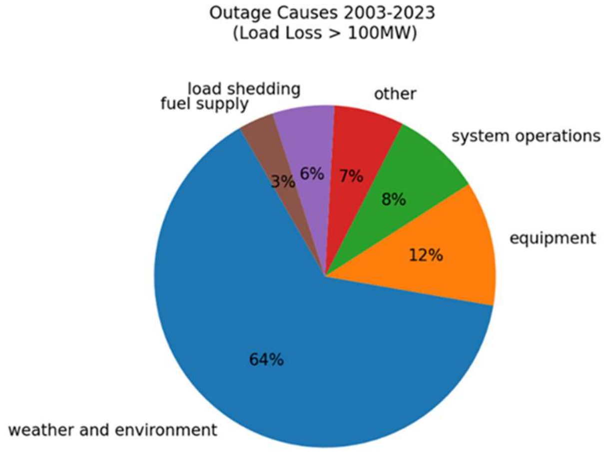


Figure 2 Major cause of power outages in the US [11]

Well known statistics of the outages in the US are shown in Figure 2 [12]. In the distribution systems, the data available from the substation recording devices may be complemented with other data as shown in Table 2. Some data (Table 2) may come from the sources that are within the utility company such as various measurements by substation recording devices and historical outage data or from outside the utility company such as weather and vegetation data [12]. This abundance of data opens an opportunity for application of ML/AI techniques to assist utility personnel, including PAC engineers in addressing outages. We illustrate the benefits from utilizing ML/AI data models in predicting outages in distribution systems in following sections.

Table 2. Sources of data for ML/AI PAC applications

	Data Class	Data Source (Measurements)	VOLUME (Data file size)	VELOCITY (Rate of use)	VERACITY (Accuracy)
VARIETY	Utility measurements	SM	120GB per day	Every 5-15 min	error <2.5%
		PMU	30GB per day	240 samples/sec	error <1%
		ICM	5GB per day	250 samples/sec	error <1%
		DFR	10MB per fault	1600 samples/sec	error <0.2%
	Weather data	Radar	612 MB/day per radar scan	Every 4-10 min	1-2 dB; $m\ s^{-1}$
		Satellite	At least 10 GB per day	Every 1-15 min	VIS<2%; IR<1-2K
		ASOS	10 MB/day per station	Every 1 min	T-1.8°F, P<1%, Wind speed - 5%, RR - 4%
		NLDN	40 MB/day	During lightning	SE < 200m, PCE <15%
		WFM	5-10 GB/day per model	15min - 12 hours	Varies by parameter
	Vegetation and Topography	TPWD EMST	2.7 GB for Texas	static	SE < 10 m
		TNRIS	300 GB for Texas	static	SE < 1 m
		LIDAR	7 GB for Harris Co.	static	HE < 1m, VE < 150 cm

SM – Smart Meter; PMU – Phasor Measurement Unit; ICM – Intelligent Condition Monitor (includes Intelligent Transformer Monitor – ITM, Circuit Breaker Condition Monitor – BCM, etc.); DFR – Digital Fault Recorder; Radar - Radio Detection and Ranging; Satellite - Geostationary and Polar-Orbiting Meteorological Spacecraft; ASOS - Automated Surface Observing System; NLDN – National Lightning Detection Network; WFM – Weather Forecast Model; TPWD EMST - Texas Parks & Wildlife Department - Ecological Mapping Systems of Texas; TNRIS - Texas Natural Resources Information System; LIDAR - Light Detection and Ranging.

In our example the focus will be on outage prediction characterized by the outage state of risk (SoR). Since such prediction can be made several hours ahead, there is an opportunity to prepare for the outages by managing mitigation measures. This approach creates a new paradigm for PAC engineers since in the current practices their focus is on capturing the fault occurrence and by locating it, they are supporting system restoration activities. The customer is typically left to their own means to cope with the outage. The outage SoR prediction, if communicated to the customers, can help them develop their own mitigation measures to reduce outage impacts as shown in Figure 3.

With the help of the ML/AI techniques, the process of correlating data from Table 2 is automated, and results shown in the SoR spatiotemporal maps are superimposed on the feeder map allowing the operators, maintenance and PAC engineers alike to manage and implement the mitigation measures, and the SoR warnings sent to the customer can be used to manage and mitigate the outage impacts.

Figure 3 The mitigation measures to reduce the outage impacts

Automated analysis of the faults in transmission system using ML/AI on synchrophasor data

Close analysis of data types and properties from Table 1 differentiates the uses and users of data. PAC engineers use data obtained from triggered instruments such as DFRs and DPRs to create reports based on analysis of analog and contact data samples. The state-of-the-art is that such data is collected automatically whenever the faults occur and then analyzed automatically [7], allowing protection engineers to characterize faults quickly, initiate appropriate repair actions by informing the repair crew where the faults are located, and inform power system operators what type of fault has occurred and whether or when the transmission lines that experienced faults may be restored. The disadvantage of using such equipment is that the occurrence of the fault triggers the recordings, which only captures a portion of the event data starting with the trigger instance and ending with fault clearance. The issue is that such data does not offer an extensive view of the disturbances reflecting the events that may be leading to a fault, including frequency excursions or disturbances caused by switching of power system equipment, including generation rejection and load shedding or transmission line switching.

Power system operators, on the other hand, get data from all substations using SCADA, but such data is somewhat limited by the recording features of RTUs. RTUs report only root mean square (RMS) analog values and contact changes, which happens by exception when certain measured values exceed thresholds. A typical scan of SCADA systems is only every two to ten seconds to help reduce the large amount of data being captured and communicated. Hence, the events occurring between the scan periods may not be captured. Besides, the RMS values without the phase angle between the recoded three-phase waveforms of voltages and currents cannot be used to analyze fault type or location. That is why SCADA data is not used to analyze faults but only allows system operators to observe switching actions caused by relays. In this section we discuss the use of ML/AI-based data analytics that focus on the analysis of streaming PMU data, which can complement the fault and disturbance analysis with system-wide analysis based on synchrophasors streamed by the PMU units typically at 30 or 60 frames per second.

We start with the *problem formulation*: Given a signal segment $\mathbf{s}(t-\Delta,t+\Delta)=[\mathbf{s}^{(1)}(t-\Delta,t+\Delta),\ldots,\mathbf{s}^{(M)}(t-\Delta,t+\Delta)]$, from multiple anonymized PMUs predict event type $y\in\{0,\ldots,C\}$ that occurred at $[t-\Delta,t+\Delta]$ by learning from scarce observations and low precision labels.

Then, we focus on the research questions to be answered as one applies ML/AI to streaming PMU data. The summary of the questions and answers is given in Figure 4 below. Our research leading to the answers is published in several paper so far [8, 9, 14, 15].

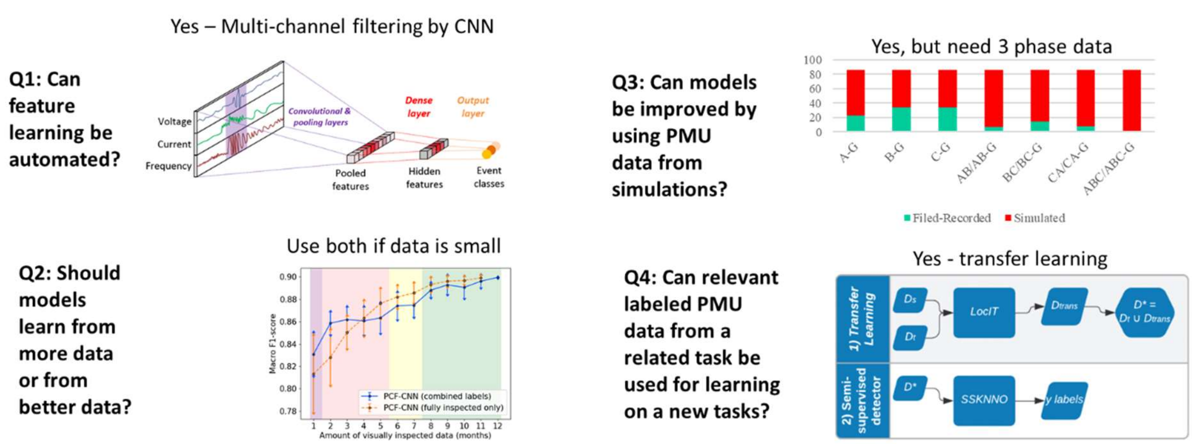


Figure 4 Research questions and answers when using ML/AI on PMU data

As it may be noted from Figure 4, some of the issues when applying ML/AI algorithms to the streaming PMU data are as follows: a) data is typically not clearly labeled, particularly regarding the time of an event occurrence, b) the data is rather noisy with lots of missing and bad data, c) some events are not equally represented in the recorded data, so additional simulations are needed to reach the balance, d) quite often, for security reasons, the topology of the power system is not provided. Besides those key issues, many other more subtle technical issues stemming from PMU data properties need to be resolved (Figure 5).

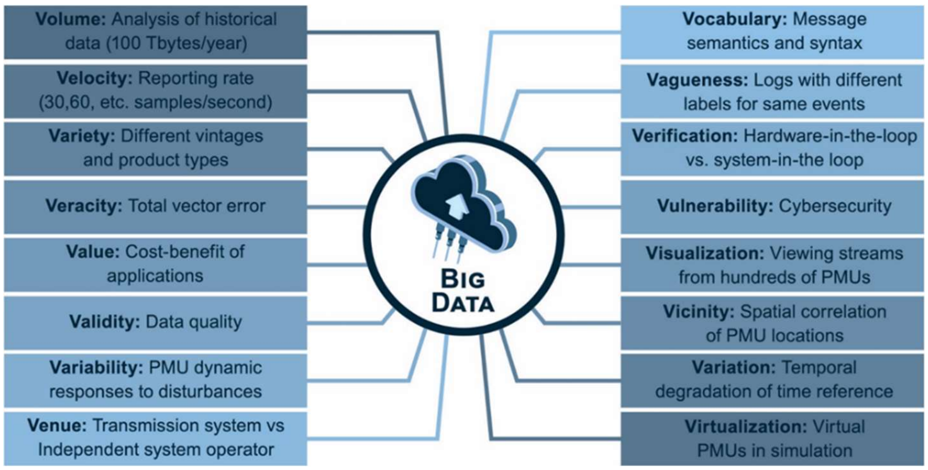


Figure 5 Property of synchrophasor data recorded by Phasor Measurement Units

While addressing all the PMU data properties depicted in Figure 5 may take considerable time and perhaps realistically may not be practical. Figure 6 shows a practical approach that yields some useful results. In this case we used field recordings from multiple sources (The US utility data provided through Pacific Northwest National Laboratory-PNNL; data provided by the French utility Réseau de Transport d'Électricité-RTE; data provided by Salt River Project-SRP) and simulated data obtained using PSCAD commercial time-domain software and an IEEE PSRC recommended fault study model). The ML/AI algorithm development requires multiple stages: collecting and processing historical recordings of synchrophasor data, then selecting the model and training it, and eventually testing the results. In this case, different ML/AI algorithms were used: a) transfer learning (TL) using transfer component analysis (TCA), b) support vector machine (SVM), and c) correlation alignment (CORAL) deep learning. Such choices came after a considerable study [8, 9, 14, 15].

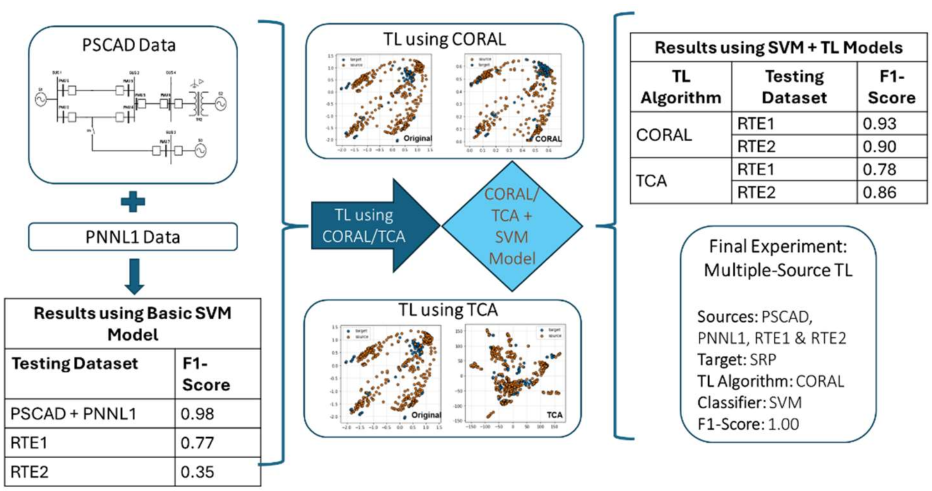


Figure 6 Training and testing stages of a fault detection/classification ML/AI algorithm

The outcome of this development was set of algorithms that may be used to automatically detect and classify faults from the incoming synchrophasor data. Since the algorithms take streaming data, any fault disturbances are captured as they occur providing PAC engineers with additional information, and PAC engineers and operators. with timely information about system wide fault events. In the case of ISO operators, this may be the only automated information they receive about the faults since the fault analysis based on digital relay and digital fault recorder data may not be shared by transmission system operators (TSOs).

Automated outage state of risk (SoR) prediction in distribution systems using ML/AI on historical outage and environmental data

This application is also unique since it allows prediction of the outage SoR leading to pre-emptive actions to reduce outage impacts. To pursue such developments, one must decide which causes of outages will be covered and what corresponding data may be used to reflect the causes. In this sense, many options are available as shown in Figure 7.

Figure 7 Causes of outages and related data sources

The left-hand side of Figure 7 shows the outage causes such as tree and animal contact, snow and lightning storms, fires, etc., while the right-hand side shows databases that contain relevant data. More details about the database sources are given in Table 2 which was discussed at the beginning of the paper. While we are showing same examples of the data sources, the details of the data features contained in the sources is quite complex depending on the type of data we may wish to extract. As an example, just the weather data may offer different features deepening on the weather sources as depicted in Figure 8.

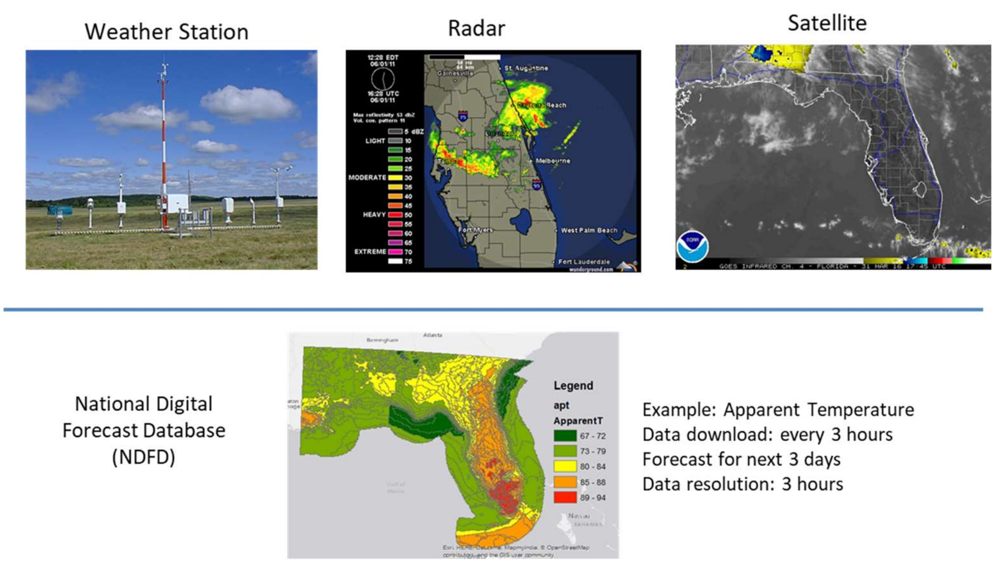


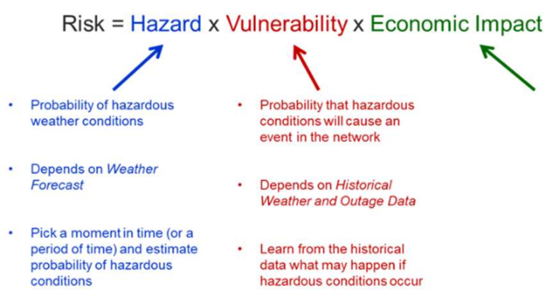
Figure 8 Sources of weather data

The choice of the weather data may be driven by availability or by the type of prediction one is trying to make. If the inclement weather is reflected only through the temperature feature, the ground weather stations may be sufficient. If the lightning data is considered, then the radar or satellite data may be used. For some more precise weather impacts, such as needed for the wind impact calculation, special treatment and processing of data may be performed to extract the relevant features [16]. In general, the weather is one of the major causes of outages, so special attention needs to be given to the selection of the weather features to be used for the study of weather impacts. If we decide to combine the impact of weather with the impacts of trees growing into the distribution feeders, then the vegetation-related data needs to be carefully considered as well. As an example, the vegetation parameters of interest and related data sources are shown in Table 3.

Table 3. Type and sources of vegetation data used for outage SoR prediction

Source	Data set name	Spatial res.	Contains
Land Boundary Information System	2004 Digital Orthographic, [11]	1 m	Imagery
National Oceanic and Atmospheric Administration	2010 C-CAP Regional Forest Fragmentation Land Cover, [12]	30 m	Forest fragmentation
Florida Division of Emergency Management	2007 St. Lucie LIDAR, [13]	Vertical 0.6 ft Horizontal 3.8 ft	LIDAR data
National Aeronautics and Space Administration	MODIS - Moderate Resolution Imaging Spectroradiometer, [14]	0.5 km	Vegetation indices
National Cooperative Soil Survey	gSSURGO, [15]	10 m	Soil data

Now, we will illustrate how the outage SoR prediction is obtained. First, we need to define the SoR, which may be done as follows:



It is apparent that the probability of the inclement weather is used for the Hazard function, the Vulnerability is calculated as the conditional probability that such weather hazard will inflict a fault, and finally the Economic Impact is the measure of the outage consequences. Once multiplied, the three components give the SoR prediction. Each of the SoR components may be represented as a Geographic Information System (GIS) map, which is then superimposed on the map showing utility feeders. The details of the outage SoR prediction process are shown in Figure 10.

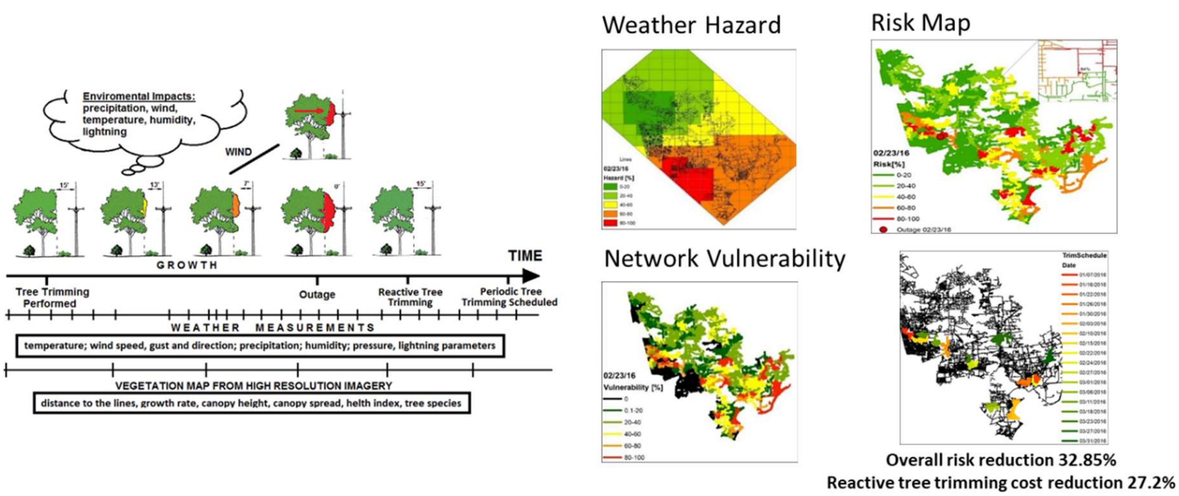


Figure 9 Outage state of risk prediction maps and mitigation outcomes

The process of developing the outage SoR prediction starts with inclusion of the weather, vegetation, soil, and many other environmental data features (left-hand side of Figure 9), combined with the historical data on outages. As a result, a hazard map and vulnerability map are generated in GIS, and then superimposed on the feeder map (right-hand side of Figure 9). The two maps are multiplied to create the outage SoR map. The economic impact may be added. Since the prediction is associated with outages caused by weather and vegetation, the parts marked in red are where the risk is highest, and vegetation needs to be trimmed to reduce the risk. The last map shows most of the areas depicted in black meaning that the risk is eliminated because of a mitigation measure of trimming trees in the areas with the highest risk.

Benefits and Requirements of ML/AI Applications to PAC engineers

Both applications have a benefit of an automated processing, which saves time and offers an ability to correlate large amounts of data well exceeding cognitive capabilities of an individual. The automation also offers consistency since it always processes the data using the same mathematical framework. The additional benefit is the impact on decision making of utility personnel about tracking power system disturbances either in near real-time (synchrophasor processing) or by predicting them. Since those are new applications, it remains to be seen how they may benefit PAC engineers. A few thoughts for the transmission applications are:

- Synchrophasors are offering system-wide assessment of the impact of disturbances, including faults. This may help PAC engineers develop a methodology for implementing adaptive relaying strategies for modifying the relaying action due to system-wide conditions.
- With an increase in the deployment of synchrophasors, additional system-wide relaying may also be implemented to offer certain system switching to alleviate the conditions leading to potentially incorrect relay tripping such as slowly developing frequency deviations or oscillations.
- In some circumstances, as the operators get informed about the impeding disturbances, they may develop some remedial actions performed manually to save the system from relay misoperation due to a possible loss of security or dependability in protective relaying operation.

In the distribution applications, several possible uses of the outage SoR prediction by PAC engineers are:

- Revisit impact of the restoration strategies on protective relaying setting, particularly in the areas of large penetration of renewable distributed energy resources since re-connecting the resources such as PV panels and electrical vehicles may create unexpected bidirectional current flows.
- Offer additional time to run the setting coordination studies to make sure that large-scale outages on important feeders do not cause any relay misoperations and implement adaptive relaying strategies as needed based on such studies.
- Analyze any safety issues as the inverter-based resources are reconnected to the grid making sure the reconnecting onto the feeders does create inadvertent feeder reenergization causing unexpected hazards for the repair and maintenance crews.

While the benefits may be explored in the future, it became apparent that the implementation of the ML/AI based solutions also imposes significant new requirements:

- Data intensive applications require new approach to data management and utilization.
- The cost of implementing such solutions is high, so the applications must create proportional return.
- This is a new practice, so acceptance and utilization by PAC engineers is still not well defined.
- The use of data requires careful cybersecurity and critical infrastructure considerations.

Conclusions

Based on the presented results and other observations about the development and implementation of the ML/AI algorithms for fault detection and classification in transmission and outage SoR prediction in distribution, the following are some conclusions:

- By its nature, the ML/AI algorithm require an extensive set of data to develop, train and test models. The PAC engineers may be able to provide some such data, but the data source vary across multiple utility groups and outside providers, so management of such sources is nontrivial.
- In the current practice, the field-recorded synchrophasor data is of a very poor quality, and the practice of not associating the data with the power system topology creates multiple challenges in implementing ML/AI algorithms, so much better utility practices of data management are needed.
- Once the weather and other data external to the utilities is used, special care has to be placed on the data cybersecurity and trustworthiness may hinder the progress in deploying such solutions in the utility environment, which may be mitigated by using a secure cloud computing.

- The PAC engineers may benefit from such applications because they become more aware of the unfolding, and future fault and outage events, allowing better understanding of relaying practices, and leading to better serving the operators and outage and asset management crews.

Acknowledgements
Former and current graduate students at Texas University have contributed to the work presented in this paper: Tatjana Dokic, Rashid Baembitov and Taif Mohamed. There are several providers of field data: CenterPoint Energy, PNNL, RTE, and SRP. The funding for the research reported in this paper is provided by CenterPoint Energy, CPS Energy, Department of Energy, and the National Science Foundation.

References:

[1] Technical Report "BPS-Connected Inverter-Based Resource-Modeling and Studies" NERC, May 2020.
[2] D. Zhang, G.M. Shafiullah, C. K. Das, and K. W. Wong "A systematic review of optimal planning and deployment of distributed generation and energy storage systems in power networks," Journal of Energy Storage, Volume 56, Part A, 2022,105937.
[3] M. Kezunovic, S. Meliopoulos, S. Venkatasubramanian, V. Vittal, "Application of Time-Synchronized Measurements in Power System Transmission Networks," Springer, 2014, ISBN 978-3-319-06218-1.
[4] H. Mohsenian-Rad "Smart Grid Sensors-Principles and Applications," Cambridge University Press, April 2022, ISBN 9781108839433
[5] M. Kezunovic, L. Xie, S. Grijalva, "The Role of Big Data in Improving Power System Operation and Protection," Bulk Power System Dynamics and Control Symposium, Rethymnon, Greece, August 2013,
[6] M. Kezunovic, Z. Obradovic, T. Dokic, B. Zhang, J. Stojanovic, P. Dehghanian, and P. -C. Chen, "Predicting Spatiotemporal Impacts of Weather on Power Systems using Big Data Science," Pedrycz, Witold, Chen, Shyi-Ming (Eds.), Springer Verlag, Data Science and Big Data: An Environment of Computational Intelligence, pp. 265–299, 2017, ISBN 978-3-319-53474-9.
[7] M. Kezunovic, P. Spasojevic, C. Fromen, D. Sevcik, "An Expert System for Transmission Substation Event Analysis," IEEE Transactions on Power Delivery, Vol. 8, No. 4, pp. 1942-1949, October 1993.
[8] M. Alqudah, M. Pavlovski, T. Dokic, M. Kezunovic, Y. Hu, Z. Obradovic, "Convolution-based Fault Detection Utilizing Timeseries Synchrophasor Data from Phasor Measurement Units," IEEE Transactions on Power Systems, 2022.
[9] A. Abdel Hai, T. Dokic, M. Pavlovski, T. Mohamed, D. Saranovic*, M. Alqudah, M. Kezunovic, Z. Obradovic, "Transfer Learning for Event Detection from PMU Measurements with Scarce Labels" IEEE Access, Vol. 9, 127420 – 127432.
[10] R. Baembitov and M. Kezunovic, "State of Risk Prediction for Management and Mitigation of Vegetation and Weather Caused Outages in Distribution Networks" IEEE Access, vol. 11, pp. 113864-113875.
[11] T. Dokic, M. Kezunovic, "Predictive Risk Management for Dynamic Tree Trimming Scheduling for Distribution Networks," IEEE Transactions on Smart Grid, Vol. 10, No. 5, pp. 4776-4785, September 2018.
[12] *Electric Disturbance Events (OE-417) Annual Summaries*, U.S. Department of Energy. [Online]. Available: https://www.oe.netl.doe.gov/OE417_annual_summary.aspx
[13] P. Dehghanian, B. Zhang, T. Dokic and M. Kezunovic, "Predictive Risk Analytics for Weather-Resilient Operation of Electric Power Systems," IEEE Transactions on Sustainable Energy, Vol. 10, No., pp. 3-15, January 2019.
[14] M. Pavlovski, M. Alqudah, T. Dokic, A. Abdel Hai, M. Kezunovic, Z. Obradovic, "Hierarchical Convolutional Neural Networks for Event Classification on PMU Data," IEEE Trans. on Instrumentation and Measurement, Vol. 70, 2021.
[15] T. Mohamed, M. Kezunovic, Y. Paudel, V. Vittal, A. Pal, D. Joshi, Menuka KC, and M. V. Venkatasubramanian, G. Torresan, "Automated Detection and Classification of Critical Power System Events using ML on PMU Data,"International conference SGSMA 2024, May 2024
[16] R. Baembitov, M. Kezunovic, K. A. Brewster and Z. Obradovic, "Incorporating Wind Modeling into Electric Grid Outage Risk Prediction and Mitigation Solution," IEEE Access, Vol 11, pp 4373-4380