Final Scientific/Technical Report

Big Data Synchrophasor Monitoring and Analytics for Resiliency Tracking (BDSMART)

WORK PERFORMED UNDER AGREEMENT

DE-OE0000913/0004

Texas A&M Engineering Experiment Station

3128 TAMU

College Station, TX, 77843-3128

Award Period of Performance: ***September 23, 2019, to June 30, 2022***

Submitted: July 31, 2022

PRINCIPAL INVESTIGATOR

Dr. Mladen Kezunovic

Texas Engineering Experiment Station

TEAM MEMBERS

Texas Engineering Experiment Station

Dr. Tatjana Djokic, Rashid Baembitov, Taif Mohamed

Temple University

Dr. Zoran Obradovic (Co-PI), Ameen Abdel Hai, Mohammad Alqudah, Hussain Otudi, Dr. Martin Pavlovski, Daniel Saranovic,

Quanta Technology

Dr. Yi Hu (Co-PI), Dr. Zheyuan Cheng, Dr. Juergen Holbach

SPONSORING PROGRAM OFFICE

U. S. Department of Energy

Office of Electricity Delivery and Energy Reliability

Via the National Energy Technology Laboratory

TABLE OF CONTENTS

Acknowledgment: This material is based upon work supported by the Department of Energy under Award Number DE-OE0000913.

Disclaimer: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS AND ABBREVIATIONS

AI	*Artificial Intelligence*
ARRA	*American Recovery and Reinvestment Act*
AUROC	*Area Under the Receiver Operating Characteristic*
AUPRC	*Area Under the Precision-Recall Curve*
CB	*CatBoost*
DOE	*Department of Energy*
EMS	*Energy Management System*
ERCOT	*Electric Reliability Council of the State of Texas*
FOA	*Funding Opportunity Announcement*
Fps	*Frames per second*
GPS	*Global Positioning System*
HPRC	*High Performance Research Cluster*
IC	*Interconnect*
kNNO & iNNE	*K-nearest Neighbor Outlier Detection*
LocIT+SSKNNO	*Localized Instance Transfer + Semi-supervised K-nearest Neighbor Anomaly Detection*
LR	*Logistic Regression*
MCC	*Matthews Correlation Coefficient*
ML	*Machine Learning*
MLP	*Multi-Layer Perceptron*
MVEE	*Minimum Volume Enclosing Ellipsoid*
PCE-CNN	*Parallel Channel Embedding Convolutional Neural Network*
PMU	*Phasor Measurement Unit*
PNNL	*Pacific Northwest National Lab*
RF	*Random forests*
ROCOF	*Rate of Change of Frequency*
SCADA	*Supervisory Control and Data Acquisition*
SC-CNN	*Single Channel Convolutional Neural Network*
SCE-CNN	*Simultaneous Channel Embedding Convolutional Neural Network*
Soft-DTW	*Soft Dynamic Time Warping Algorithm SVM-Support vector machine*
TEES	*Texas A&M Engineering Experiment Station*

I. Executive Summary

This report contains key findings from a project titled Big Data Synchrophasor Monitoring and Analytics for Resiliency Tracking (BDSMART), which was carried out through a collaborative effort of a team of researchers from Texas A&M Engineering Experiment Station, Temple University, and Quanta Technology, LLC. The in-kind support came from OSIsoft (acquired by AVEVA), which provided their PI Historian software to demonstrate the use case of streaming PMU data.

The first section of the report describes the project goals and objectives related to the development of Machine Learning (ML) models capable of detecting and classifying events by processing phasor measurements captured in the field by Phasor Measurement Units (PMUs). The data for this study was contributed by the utilities/ISOs from the Western and Eastern interconnects and ERCOT, further referred to as Interconnect B (IC B), Interconnect A (IC A), and Interconnect C (IC C), respectively. The approach that the BDSMART Research Team proposed and the key research tasks defined by the team are outlined in this section.

The next section describes the technical approach. We first discuss the data constraints related to the PMU measurements and data interpretation constraints imposed by the data contributors. They provided neither the topological information of the grid nor PMU placement locations and captured recorded data at very few locations in the system with the reporting rate of either 30 or 60 fps. The recordings are mostly positive sequence voltage, frequency, and ROCOF, and in some limited cases, three-phase voltages and currents. We then reflect on the bad data issues that stem from poor recording practices and vague definitions of the PMU status bits to supposedly be used for bad data identification. Finally, the data discovery points to imprecise time stamps with incomplete event start/end time, as well as inconsistent and incomplete event labeling, which combined make the implementation of the data models using supervising learning quite challenging. Following the data discovery study, we hypothesize that because the IC B data has the most complete labels, we should focus our model development on that data and then test it on data from other interconnects. We also define the common metrics used to evaluate the results from the ML algorithm tests. We concluded this section by summarizing the common ML models we used and explaining how we implemented and tested them. The issues from this section are expanded in the Training Dataset Report from this project.

The final section of this report deals with the accomplishments and conclusions. As the accomplishments, we formulate the problem we are solving and what is achieved by solving the problem. We then reflect on each of the analytics tools we developed and point out the performance of each tool when applied to solving the mentioned problems. We reference this work for further details to the papers we published on each tool. In the conclusions, we give recommendations on how to improve future PMU recording practices to facilitate the ML algorithm implementation and guidance for the future standardization work aimed at clarifying the ambiguities associated with the PMU status bits. We finally list future tasks that can bring about further improvements in the proposed algorithms. The issues from this section are expanded in the Training, and Test Dataset Report filed at the project completion date.

II. Objectives

This section discusses the goals and objectives stated in FOA 1861 and then reflects on how the proposers have defined the objectives of their project activities together with the software implementation framework, project scope, and tasks.

II-a. The Goal of FOA 1861

According to FOA 1861, "The goal of this funding opportunity announcement (FOA) is to explore the use of big data, artificial intelligence (AI), and machine learning technology and tools on PMU data to identify and improve existing knowledge, and to discover new insights and tools for better grid operation and management. Applicants selected for the award will receive pre-packaged datasets assembled exclusively for their use executing awards resulting from this FOA. Applicants selected for the award will be asked to address specific questions and research areas regarding the data. Applicants selected for the award will publicly present their analytical results at a DOE-sponsored event to inform stakeholders in the electricity sector who develop and use analytical and decision-making tools on PMU and other power system data."

II-b. The Objectives of FOA 1861

The study assumed that synchrophasor field data collected in the Eastern and Western interconnect and ERCOT would be shared by PNNL with the contracted teams. PNNL received the data from utilities/ISOs in the mentioned interconnects – then organized it into two datasets. (a) Training dataset covering approximately ¾ of the recordings from two years of data. (b) Test dataset that was obtained by cutting out sections of the overall dataset and packaging them for data model testing purposes.

The FOA 1861 further stipulated that "Applicants selected for the award will analyze the data provided looking for key patterns and insights, potentially including but not limited to several of the following purposes:

1. Identify key events within each interconnection-specific dataset
2. Identify unusual or anomalous events and patterns that did not appear in the state estimator and event log.
3. Catalog of the "signatures" (identifying patterns), including events (e.g., faults, tree contacts, frequency excursions/oscillations, fault-induced delayed voltage recovery (FIDVR), wind ramps, clouds over photovoltaic (PV), cyber attacks, time-signal problems, or oscillations), harmonics, and asset problems (e.g., deteriorating equipment, erroneous machine settings) that can be used for reliable identification and diagnosis of different events and conditions on the grid.
4. Identify precursor conditions that warn about events that occurred after the associated conditions.
5. Identify patterns that reveal the condition of power system equipment and any insights into equipment modeling or predictive or condition-based maintenance on assets, including the PMUs themselves.
6. Identify apparent ground-induced currents (GIC) relating to geomagnetic disturbances, along with any potential impacts of GICs on asset conditions.
7. Predict the performance of power plants and other assets.

8. Identify actual load events and patterns, including potential detection of power sources such as distributed generation or storage located on or below the distribution system.
9. Identify factors that can improve wind integration and solar integration.
10. Identify factors that reflect weather and seasonality without using site-specific weather data.
11. Identify anomalies that may reflect cyber security issues and events rather than grid performance."

II-c. The Approach Proposed by the Research Team

In response to FOA 1861, Texas A&M Engineering Experiment Station (TEES) submitted a proposal in collaboration with Temple University, Quanta Technology, and OSIsoft (non-funded partner) titled "Big Data Synchrophasor Monitoring and Analytics for Resiliency Tracking" (BDSMART).

The following were detailed objectives of the BDSMART project:

- Implement software embedded in a commercial OSIsoft platform that will allow automated analysis, classification, and prediction, and develop techniques to analyze recorded PMU data files to differentiate the ones that are well defined and correlated with particular power system events from the ones that need further assessment to better understand the causes
- Develop a precise and fast event classification algorithm capable of taking PMU files as inputs and identifying the potential cause of disturbance in real-time as outputs
- Develop a prediction framework for estimating the probability of multiple types of events in the near future using disturbance precursors

To meet the objectives, we proposed to follow a three-step software implementation framework:

1. Analysis: To create a data analytics platform capable of integrating any data source of interest into the database. This will ensure an environment that can support ingestion, curation, feature extraction, and visualization of large amounts of historical data and real-time data streams.
2. Classification: The purpose of this step was to provide event type labels for each synchrophasor recorded disturbance. The focus will be on assuring – (1) postmortem identification of every historical disturbance class and (2) real-time automated classification of event types using previously acquired knowledge about classes.
3. Prediction: This step was to determine the probability of various events in the near future after the occurrence of one or more precursor events recorded by the synchrophasor system.

To deal with the research tasks, we proposed the following set of *project tasks*:

- Task 1.0 - Project Management and Planning (Lead: TEES)
- Task 2.0 - Raw Dataset Clarification and Initial Results Evaluation (Lead: Quanta Technology)
- Task 3.0 Event Analysis (Lead: TEES)
- Task 4.0 Event Labeling/Categorization (Lead: Temple University)
- Task 5.0 Event Prediction Based on Precursor Conditions (Leads: TEES and Temple)

- Task 6.0 – Evaluation (Lead: Quanta Technology)

III. Technical Approach

The technical approach was driven by the data constraints discovered in the first year of the project – followed by the hypothesis – and various experimental studies with a variety of data model approaches tailored to the specific data properties – discovered by inspecting the data provided by PNNL and contributed by utilities/ISOs.

III-a. Data Constraints

The data constraints driving the technical approach relate to the quality of PMU measurements, data wrangling, and other considerations.

III-a.1. PMU Measurements

*No information on power system topology***.** The datasets were not referenced to any particular power system topology, so it is unclear where the data is collected (buses, voltage levels, specific grids, etc.). We did receive information on what interconnect the data is coming from, but everything beyond that had to be speculated or estimated. This created an inability to develop any physical system models that could be used to correlate the measurements or simulate the events, except for fault simulations, which were feasible since they could be performed on a small size generic model and still be reasonably accurate since fault waveforms have limited correlation with the voltages and currents in the parts of the system that are distant from the fault.

Unknown PMU locations. Since the PMU location is unknown, and events are not labeled with the specific location in a power grid, we do not know how far from the occurrence of an event a PMU is located. For the local events such as faults, some electrical signal properties caused by event occurrence are far less prominent measured at a distance from the event occurrence. A typical example is recording the changes in the current and voltages at the PMUs not close to the faulted line, which is very hard to differentiate from normal operation or some other unrelated events. Not knowing the PMU locations creates a challenge in correlating such measurements from multiple PMUs to determine whether the measurements can be used to detect the same event from multiple local measurements.

Excessive measurement sparsity. The number of buses in the Eastern Interconnect (IC C), Western Interconnect (IC B), and ERCOT (IC A) is approximately 70,000, 20,000, and 4,500, respectively. We are dealing with only 188, 43, and 212 PMUs in the mentioned interconnects, respectively, which makes the measurement sparsity problem obvious. PMU data provided by DOE may have come from more densely located PMUs within several smaller areas in an interconnect if data providers are from a few data-contributing utilities in that interconnection, which at the scale of the whole interconnect, makes the sparsity uneven over the geographical area of the interconnect.

Incomplete measurement signal set. A power system typically has three phases fully characterized by three voltages, three currents, and aggregate frequency. In some PMUs (very few), all such signals are measured and sent, but in many others, only positive sequence voltage (sum of all three phases), frequency, and rate of change of frequency (ROCOF) are sent. This PMU output configuration is decided by the users. In general, PMUs always compute synchrophasor measurements for all three phases. These measurements are then used to calculate the positive, negative, and zero sequence Synchrophasors. Which signals to send out is entirely up to the user to decide by selecting appropriate PMU output configuration settings. Some events (i.e., faults) require three-phase information for full characterization of the event, so the use of positive sequence voltages and currents does not allow for fully characterizing such events.

Variable phasor reporting rate. The phasor reporting rate is NOT the same as the sampling rate. The sampling rate is associated with the waveform sampling for analog to digital (A to D) conversion, which typically could be in the order of several hundreds to millions of samples per second. The phasors are calculated based on a data window containing many waveform samples and reported at 30 (or more) frames per second (fps). In the DOE-provided data, some were reported at 30 fps and some at 60 fps. This created a need to re-sample some data to make the reporting rate consistent across a dataset used for model development.

Data files of large size and restraining format. The size of the historical data received for the Eastern and Western Interconnect and ERCOT is 16.3TB, 6.87TB, and 4.1TB, respectively. The data has been imported and kept in the parquet format on a high-performance research computing (HPRC) cluster at Texas A&M University. This format turned out not to be the most efficient format to run the experiments, and sometimes it took days to complete specific data model evaluation tasks making the experimental part of the project rather tedious, time-consuming, and computationally inefficient.

III-a.2. Data Wrangling

Missing information. Criteria used to decide which events are included in provided PMU datasets are unknown to us. For the test datasets provided from different interconnects, it is an expectation that our study would find out how many events the PMUs have captured. But even for the training datasets, the number of events in these datasets is more than the number of events in the event logs that DOE provided. This is because (1) the DOE dataset did not provide event logs for any other events in the IC A but the generator tripping events and did not include all events of the original event log in the updated event log with improved labeling for IC C; (2) Events that do not involve major switching operations likely are not logged by utilities; and (3) A utility that provides the data will not log events in its neighboring utilities even though its PMUs will be able to “see” those events. This makes it difficult to evaluate the model accuracy since there is no reference to an actual set of events that could be detected to compare the results to.

Bad and imprecise data. By performing manual inspection and observing PMU data quality bit indicators and/or thresholding, one can conclude that there is a large percentage of bad data. Since PMU types are unknown, the measurement accuracy may be imprecise for given events, and bad data is not equally present in a dataset and among PMUs. This created an issue with implementing automated bad data detection approaches since the information about the measurement system is not known, and the interpretation of the PMU status bit as indicators that are supposed to recognize errors is not possible.

Complex feature engineering. The electrical system exhibits dynamic signals caused by the various events that are in different frequency ranges from a fraction of a Hz to 60 Hz (nominal) and many multiples of the fundamental frequency. Many time-domain and frequency domain features may be extracted, and a limited number of features is used in our study. Guidance of what features are most prominent for which type of signal and the related event was not a trivial task, so some attempts at feature engineering that looked promising early on have proven difficult to utilize in some other cases.

Vague labeling. The labels in the provided event logs are assigned manually and hence may be imprecise, incomplete, or totally wrong. They are also not very accurately time-stamped. In some event instances, they are simply missing. Several practices led to the problems we see in the datasets and the accompanying event logs, and such practices are related to the system or process used to characterize the events, which in most instances is the Energy Management Systems (EMS) Supervisory Control and Data Acquisition (SCADA) rather than the Synchrophasor system. This created a situation where a less accurate system (SCADA) is used to characterize a more accurate system (Synchrophasors), which is opposite to the practical and theoretical requirements used in the measurement and instrumentation industry.

Imprecise time stamps. Utilities generally log an event when switching operations are involved with major components (i.e., lines, transformers, generators, etc.) in the system. EMS/SCADA databases were likely the main source of records of such events as all breaker status changes are logged there. This could also explain why the event time is not very accurate, as EMS/SCADA does not use an accurate time clock, and the time stamp from such databases can be imprecise. Oscillation events are recent addition that may not involve any switching actions.

Inconsistent labeling practice. Most likely, the three interconnects may not have any mandatory requirements for labeling the events uniformly and consistently. This may have resulted in each utility labeling the events in their own ways. At least we know this was not consistent between Eastern and Western interconnects. It is possible that this may not be uniform and consistent even within one interconnect (especially for the Eastern interconnect, which has multiple reliability coordinators.).

Misinterpretation of events. Each utility only logs events involving major switching operations within its own control area, but their PMUs should be able to "see" similar events occurring in their neighboring utilities. Since many more PMUs have been installed and operating in these interconnects today than what was used to provide data for this study, it means that not all the utilities have provided their PMU data. This could lead to the inclusion of the events that can be "seen" in the datasets but not correspond to events in the provided event logs if the events occurred outside the areas of utilities that provided the data.

Limited analysis practice to confirm the historical events. One may assume utilities always do a detailed post-event analysis for every event before they label the event. However, they may not be able to always positively determine the causes of the event for various reasons, such as lack of high resolution (e.g., point-on-wave) historical data preceding the event. Without any reference to how the events are analyzed to assign labels, it is not clear whether the labels are consistently assigned.

III-a.3. Other Considerations

Unknown PMU compliance with the standards versions. PMUs are calculating phasors based on samples of a time-domain electrical signal. By the nature of the measurement model, many components of the time-domain signal are filtered out. PMUs are filtering the time-domain signal differently based on the measurement performance requirements. In the 2005 PMU standard, there were two levels (0 and 1) of measurement performance requirements, which were changed to P and M classes in the 2011 standard. Many PMUs deployed earlier and even during DOE's ARRA projects (2010-2012) are based on the 2005 standard. Not knowing the type, model, and vintage of PMUs being used and which PMU standard version they are complied with created difficulties in making assumptions about the measurement model and interpreting the exact meaning of STAT bits.

Limited synthetic model availability. For fault studies, we used a small size generic model of a power system that does not represent many typical components and is incapable of accurately representing various power system-wide disturbances. Despite its lack of generality, it has some value for fault studies. We used it to generate PMU data with simulated fault related events that we then used to train the data models in addition to train these models with field recorded data.

Unknown event causality. The dataset descriptors provided neither causality of the events nor their interdependencies. We also do not know whether or what automatic equipment or operator action was taken in response to an event or sequence of events. In many instances, we do not know the event length (time window) either. Lacking the information that suggests causality of the recorded events or any indication that correlates the recorded events to the power system topology and PMU placements, the ability to develop the causality models was significantly constrained. We reached out to an additional option of determining the causality by correlating the recorded events to the severe weather events in the region when the recordings were made and were able to discover causalities between bad weather occurrences and subsequent faults.

*Variable data window***.** Visual inspection of events confirmed that different types of events have different durations. Certain events, such as local line fault events, are typically short, whereas other system-wide events, such as frequency oscillation events, can last significantly longer. Event durations range from <1 sec up to several hours. Consequently, such a variety in the event durations considerably hinders their detection with a universal data model. Furthermore, the start and end times of events in some instances are missing in the event log, making the problem of event detection and data window-size selection even more difficult. Therefore, various window sizes are required to address the issue of different event durations. This problem is compounded by the fact that large window sizes might contain overlapping events.

III-b. Hypothesis

The hypothesis was developed based on the discovery that IC B data has the most complete labels, and its data quality was reasonably well understood after an extensive study of the data files. The hypothesis discussion has two aspects, namely the hypothesis formulation and the hypothesis evaluation metrics and process.

III-b.1. Hypothesis Formulation

After inspecting the data, we realized that the most complete dataset was from IC B. We also discovered that the data quality shows outliers, which in more controlled data recording practices should be eliminated. As a result, we postulated the following assumptions that can contribute to the hypothesis formulation:

- If it is possible to develop models using labeled data from the best dataset, the hypothesis is that such models will be quite useful since the practices from the labeling can be propagated to other interconnects to achieve better results in those interconnects
- Since we had to deal with imprecise and inaccurate labels, we assumed that a domain expert can inspect the labels visually to improve them, so we hypothesized that the improved labels through such a process would yield better results than without such a process being followed
- We recognized that the data window that captures and differentiates various events is a crucial component of the model development, so we hypothesized that using a sliding data window would lead to improved results

- We also noticed that the data contained records of power system faults, which we assumed are hard to characterize due to the fact that the PMUs that captured the waveforms were far from the fault occurrence, so we made a hypothesis that supplementing the training process with the fault data from a synthetic system that have virtual PMUs placed close to the location of simulated faults may improve the results obtained by the models aimed at fault detection and classification
- We assumed that the events in the various power systems might be characterized by some generic waveform signatures – so we also hypothesized that transfer learning model application from one interconnect with good labels to another with less labeled data will enable accurate event detection and classification in interconnecting with poorer labels

III-b-2. Evaluation Metrics and Process

The following evaluation metrics were used to assess and quantify the performance of the ML algorithms: (1) The Area Under the Receiver Operating Characteristic (AUROC), (2) The Area Under the Precision-Recall curve (AUPRC), (3) Precision, (4) Recall, (5) F1-score, (6) Matthews Correlation Coefficient (MCC), and (7) Accuracy. Results reported were obtained by computing the aforementioned evaluation metrics based on categorizations and/or detections of time windows. Event detection algorithms were trained using a portion of the training dataset of IC B and IC C, and tested on the remaining part of these datasets. We also refer to test data as a subset of the training data of IC B, IC C, and IC A used to evaluate the algorithms. The Test instances were not seen by the detection algorithms while training. Hence, the model was trained and tested on different time intervals of recorded data. Labels of predicted time windows were compared to the ground truth labels to quantify the performance of the algorithms.

In the process of data model development, we used different methods of splitting the data into training and test subsets. Some of our event detection algorithms were trained on time windows from IC B data from 2016 to predict future instances from 2017. In this setting, the model learned the relationships based on instances collected from 2016 without incorporating any instances from 2017 during learning to assess how well the algorithm would perform on future instances. Other event detection algorithms used the conventional way of splitting the training and test datasets, which is ~70% training and 30% testing. In this setting, 70% of the data were randomly selected for training, and the remaining 30% were used to evaluate the algorithms. Furthermore, Stratified K-folds Cross-validation was also used. Similarly, this setting learns on multiple proportions of training data instances and used different proportions of test data while preserving the percentage of instances of each class. All settings ensure that there were no information leaks between training and test instances for an accurate model evaluation.

The developed algorithms were also tested on the test data provided by PNNL on behalf of DOE. We had no accesss to testing event logs and therefore we applied the developed algorithms to classify or detect events, and detected events were ordered by their probabilities, indicating how likely is a time window to contain an abnormal behavior or to belong to a certain event type. Then a domain expert in our team manually inspected the results of each algorithm to assess their performances.

III-c. Data Model Approaches

The data model approaches relate to the algorithm selection and evaluation process. Since we decided to demonstrate the algorithm performance primarily on the dataset that has the most reliable and descriptive labels (IC B), the comments and results are primarily related to the model performance on the IC B data, even though we also tested some models on the IC C and IC A data to show the variation in performance when data with poorer labels and worse quality are used.

III-c.1. Algorithm Selection

Several traditional supervised learning methods (RF, CB, SVM, LR, and MLP) for line faults, transformer faults, and frequency events detection that rely on knowledge-based feature selection were developed. This is compared to automated representation learning based on several deep-learning-based alternatives. To mitigate the challenge of learning infrequent fault types, simulated data was also used together with field recordings. A summary of developed algorithms is presented in Figure 1. Table 1. presents an overview of algorithm characteristics that illustrates the setup of each model.

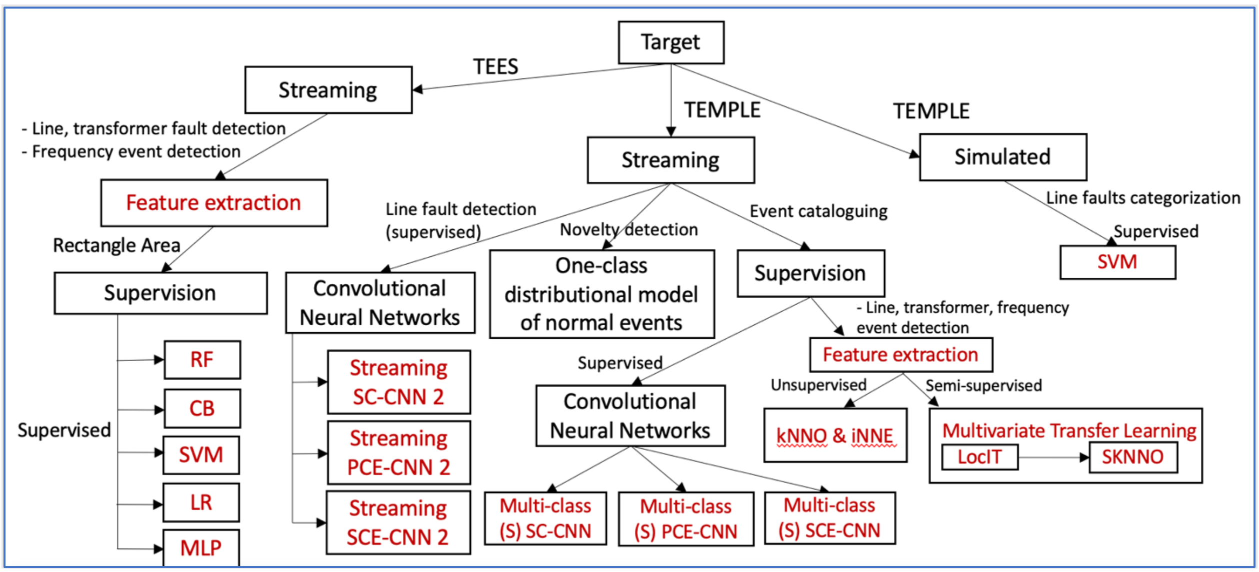

Figure 1. A summary of event detection/cataloging algorithms used in this study.

Table 1. An Overview of the Algorithm Characteristics

Subtask	Algorithm	Characteristics						
		Input signal(s)	Goal	Supervision	Feature extraction	Window size	Labels	Int.
4.1 and 4.2	Supervised Classification (RF, CB, SVM, MLP, LR) based on RA *	Voltage, frequency	Event Detection (line and transformer faults, fundamental frequency events)	Supervised	Rectangle Area	3 min, 1 min, 30 sec, 10 sec, 5 sec, 2 sec	Event Log/ Inspected by a domain expert	A, B, C
5.1 and 5.2	SC-CNN *	Voltage or current or frequency	Line fault detection	Supervised	Soft-DTW	1 min	Inspected by a domain expert	A, B, C
	PCE-CNN *	Voltage, current and frequency						
	SCE-CNN *							
	One-class distributional model of normal events*	Voltage, current and frequency	Novelty detection	Supervised	Soft-DTW	1 min		
	Multi-class SC-CNN*	Voltage or current or frequency	Event cataloging	Supervised	Euclidean Barycenter + Downsampling for Visual Representation	window: main: 55 sec; sub: 30		
	Multi-class PCE-CNN*	Voltage, current and frequency						
	Multi-class SCE-CNN*							
	LocIT + SKNNO *	Voltage and frequency	Frequency, line, and transformer events detection	Semi-supervised	Rectangle Area	1 min, 30 sec, 10 sec, 5 sec, 2 sec		
	kNNO & iNNE*	Voltage and frequency		Unsupervised				
	3-phases (SVM)	Voltage	Line fault cataloging	Supervised	3-phase features	2 sec	Log of simulated events	Simulated, B

III-c-2. Algorithm Evaluation

The model evaluation was done in several settings, some prescribed by PNNL and some introduced by the project team:

- *Model Training Using Field-recorded Data.* PNNL provided a dataset called the Training Dataset, which contained the first six weeks segments of PMU recordings from each eight-week chunk of year 2016-2017, depending on the interconnect in question. The team was asked to train the algorithm using such a dataset and report on the findings. We used year 2016 part of the Training Dataset to develop models and year 2017 part is used to evaluate the accuracy. The visual inspection on the given training dataset has resulted in a number of observed patterns and signatures. Based on the conclusions of visual inspection, a one-minute data window for event detection and classification was selected. Various machine learning types were used for fast event and fast frequency detection. The detailed results were provided in the Training Dataset Report submitted to the DOE website on June 30, 2022.

- Model Testing Using Field-recorded Data. Another dataset provided by PNNL was called the Test Dataset, which consisted of the last two weeks segments taken from each eight-week chunk of *PMU recordings from 2016-2017* . The team was asked to test algorithms using the Test Dataset, as well as a combination of the Test Dataset and Training Dataset, which combined was a complete dataset of the PMU recordings across all three interconnects. The results were provided in a required Training, and Test Dataset Report submitted to the DOE website on June 30, 2022. Each of the published paper listed in Appendix A contains an aspect of the model testing results, which are also discussed in the Accomplishments and Conclusions section of the report.
- Model Testing Using Simulated Synthetic Network Data. For the development of the data models aimed at detecting and classifying power system faults, which are local events, the team explored the use of simulated data from a synthetic network, which has the advantage of being precisely labeled since the simulations are set up by the experiment staff. We reported the results of such model training when compared to the model training performed using field recorded data. This was done since the PMUs were located away from the event occurrence locations, so the waveform signatures were not prominent enough to fully differentiate faults events from normal events and properly classify them, which created many challenges for model development. Further details are discussed in the following section of this report.

IV. Accomplishments and Conclusions

The Project accomplishments are summarized through a discussion of the various publicly disseminated results and conclusions through the recommendations that came out of completing various project tasks.

IV-a. Accomplishments

The accomplishment discussion is divided into general accomplishments and specific analytical tool developments.

IV-a.1. General accomplishments

The general accomplishment may be summarized as follows:

- Developed a supervised learning method for local and system-wide event detection in power grids using sparsely placed PMUs

  Problem: Develop a scalable, automated event detection system that doesn't rely on the extensive manual study of data and feature engineering. Utilize a sparse set of PMUs that don't necessarily cover all geo areas. Handle noisy data and unknown event locations.

  Accomplishments: Automated data preprocessing steps and three CNN-based detection models were introduced. Data from western interconnection were used (2016 for training and 2017 for testing). Robust event detection is achieved in multiple settings, but multi-channel hierarchical CNNs outperformed alternatives. Curating event logs leads to increased detection accuracy. Fully inspecting at least two months of data is suggested.

- Developed a line fault, frequency, and transformer event detection method based on transfer learning techniques

  Problem: Detect events based on minimal labeled time windows by leveraging related labeled instances from another domain without relying on event logs of PMU data.

  Accomplishments: The transfer learning method yielded ~13% improvement in AUROC when compared to supervised learning algorithms based on only 20 labeled time windows incorporated. 2-seconds time window yielded an approximately 7% increase in AUROC compared with 1-minute windows.

- Developed line faults classification using ML on three-phase voltages

  Problem: Classify line faults when field recordings have insufficient observations of a certain type (e.g., PP, PPG,3P, and 3P-G faults).

  Accomplishments: A classification model trained on integrated simulations and field-recorded data resulted in 98.5% accuracy. This significantly improved over 87.17% accuracy obtained by relying on the field-recorded data alone.

- Developed voltage level-aware CNN classifier based on sliding window technique

  Problem: Classify events as normal, line, and frequency events using PMU measurements split into voltage levels (134kV, 240kV, 345kV, 500kV).

Accomplishment: A 30s sliding window is subsampled from the preprocessed data, and those sub-windows are used as CNN classifier inputs. A split into voltage levels is beneficial when classifying signals originating from multiple sub-networks and sources of different scales.

IV-a.2. Analytical tools

Several analytical tools have been developed and reported in various publications as listed in the appendix:

R. Baembitov, T. Dokic, M. Kezunovic and Z. Obradovic, "Fast Extraction and Characterization of Fundamental Frequency Events from a Large PMU Dataset Using Big Data Analytics," Proc. 54th IEEE Hawaii International Conference on System Science (HICSS-54), Hawaii, USA, January 2021. pp. 3195-3204, doi: 10.24251/HICSS.2021.389.

Algorithm: Fast extraction and characterization of fundamental frequency events. Catboost and Random Forest classifiers were used for extraction, and Minimum Volume Enclosing Ellipsoid (MVEE) was used for characterization.

Problem: Enable the fast extraction of fundamental frequency events from extremely large historical PMU datasets. Next, develop an algorithm that accurately characterizes the event duration for the extracted events.

Performance: Based on the Area Under the Receiver Operating Curve (AUROC) Catboost performed best at 98% accuracy. MVEE successfully characterized 93.72% of these events, revealing an average duration of 9.93 seconds.

M. K. Alqudah, M. Pavlovski, T. Dokic, M. Kezunovic, Y. Hu and Z. Obradovic, "Fault Detection Utilizing Convolution Neural Network on Timeseries Synchrophasor Data From Phasor Measurement Units," in IEEE Transactions on Power Systems, doi: 10.1109/TPWRS.2021.3135336.

Algorithm: Convolutional Neural Network end-to-end supervised learning tool with automated feature learning trained from sparse PMU data

Problem: The model is developed for the purpose of fault detection. In addition, the model addresses the issue of automated feature learning from multiple PMU signals by learning high-order features without the need for manual feature engineering, which can further be used for downstream tasks such as fault detection.

Performance: On western interconnection data, results showed robust detection accuracies (AUC of 83%), when trained on 2016 and evaluated on 2017 data.

M. Pavlovski, M. Alqudah, T. Dokic, A. A. Hai, M. Kezunovic and Z. Obradovic, "Hierarchical Convolutional Neural Networks for Event Classification on PMU Measurements," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-13, 2021, Art no. 2514813, doi: 10.1109/TIM.2021.3115583.

Algorithm: Hierarchical Convolutional Neural Network tool for multi-class event classification utilizing automated feature learning

Problem: This model aims to discover the effects of the quality of labels on multi-class classification models using sparse PMUs.

Performance: Using 2016 western interconnection data for training and 2017 for evaluation, accuracy of 89% to 94% is obtained using this tool. The study showed that by manually inspecting 2 months of data, satisfactory classification results could be obtained, and inspecting 8 or more months of data can yield improved detection results, but at the cost of significantly increased labeling efforts.

Abdel Hai, T. Dokic, M. Pavlovski, M. Alqudah, M. Kezunovic, Z. Obradovic, "Transfer Learning for Event Detection from PMU Measurements with Scarce Labels," IEEE Access, Vol. 9, 127420–127432, September 10 2021, doi: 10.1109/ACCESS.2021.3111727.

Algorithm: Event detection tool based on the Transfer Learning technique

Problem: The tool was developed to address the limitations of automating event labeling in high-volume PMU measurements. Existing historical event logs created manually do not correlate well with the corresponding PMU measurements due to scarce and temporally imprecise labels. Trying to overcome this problem by extending event logs to a complete set of labeled events is very costly and often infeasible. Thus, this method reduces the need for additional labeling by leveraging related labeled data instances available from a small number of well-defined event detection tasks.

Performance: Results show that this method can significantly improve event detection based on PMU measurements when extensive labeling is costly or infeasible to obtain - with only 20 labeled time windows, this method achieved 0.93 AUROC.

H. Otudi, T. Dokic, T. Mohamed, Y. Hu, M. Kezunovic, Z. Obradovic, "Line Faults Classification Using Machine Learning on Three Phases Voltages Extracted from Large Dataset of PMU Measurements," Proc. 55th IEEE Hawaii International Conference on System Science (HICSS-55), Hawaii, USA, January 2022. DOI: 10.24251/HICSS.2022.425.

Algorithm: Three-Phase Line Fault Classification algorithm is developed to supplement field-recorded PMU data with simulated data

Problem: The model is used to supplement field-recorded PMU data with simulated data where the number of certain types of events is insufficiently observed in the field recordings. Moreover, the same algorithm is applicable when some fault types are difficult to distinguish in field recording data.

Performance: The results after training this model with the combined dataset showed a classification accuracy of 98.58%, which is a significant improvement compared to the 86.87% to 87.17% accuracy obtained by relying on the field-recorded dataset alone.

Z. Cheng, Y. Hu, Z. Obradovic, M. Kezunovic, "Using Synchrophasor Status Word as Data Quality Indicator: What to Expect in the Field?", IEEE Smart Grid Synchronized Measurement and Analytics Conference, SGSMA 2022, Split, Croatia, May 2022.

Algorithm: A point-by-point data quality issue, and STAT bit flags identification and counting algorithm. At each time point, the data quality issues such as missing data, unreasonable values, erroneous time tags, and many other issues are identified and counted, and the STAT bits, if they are not zero, are interpreted and counted for every PMU.

Problem: Field recorded PMU data contain various types of data quality issues that could impact AI/ML algorithms. If data quality issues are not thoroughly identified and counted, their potential impact to AI/ML algorithms cannot be properly evaluated for taking appropriate actions to mitigate their adverse impact.

Performance: All data values and STAT bits in the DOE-provided datasets have been scanned, and all identified data quality issues are counted. It was determined that the STAT bits are not a reliable indicator due to inconsistencies in bit setting and mismatch between the STAT bits and actual data quality issues of the data. The results have helped successful AI/ML algorithm development and testing in this project.

T. Dokic, et. all, "A Single-Feature Machine Learning Method for Detecting Multiple Types of Events from PMU Data," IEEE Smart Grid Synchronized Measurement and Analytics Conference, SGSMA 2022, Split, Croatia, May 2022.

Algorithm: A single-feature event detection algorithm. The single feature is called rectangle area (RA), extracted using the maximum and minimum values of positive sequence voltage and frequency.

Problem: Historical PMU measurements form extremely large datasets that usually contain data quality issues such as missing, duplicated, or out-of-range data. The challenge is developing a machine learning algorithm that can perform well using this data and quickly extract events without the need for several data preprocessing steps.

Performance: The Random Forest classifier performed best using the RA feature. The maximum accuracy of event detection reached initially was 0.931. The accuracy increased to 0.991 after improving the provided labels further through visual inspection.

Various aspects of the application of the models developed to detect and classify faults were published in the following papers:

- M. Kezunovic, et al. "Automated System-wide Event Detection and Classification Using Machine Learning on Synchrophasor Data," CIGRE General Session, Aug. 28-Sept 2, Paris, 2022.
- M. Kezunovic, et. al. "Use of Machine Learning on PMU Data for Transmission System Fault Analysis," CIGRE General Session, Aug. 28-Sept 2, Paris, 2022.

IV-b. Conclusions

The overall conclusions fall into three categories of recommendations related to improved PMU recording practices, future standardization work, and Opportunities for further data model developments and improvements.

IV-b.1. Improved PMU recording and labeling practices:

The recommendations are as follows:

- Many data quality issues result from practices where the quality of field data is not checked regularly. Implementing a method for continuous data quality checks will eliminate this nuisance that deeply affects the development and use of the data models
- Proper labeling can significantly improve the data model selection and accuracy. Data labeling should be done not only based on SCADA data but also based on data from other recording systems, including GPS time-stamps
- The lack of the power system topology and PMU placement information significantly impairs the data model development and selection. It is essential to synergistically combine

machine learning models with power system domain knowledge since data-based models, as powerful as they are, will not be sufficient

IV-b.2. Future standardization work:

The recommendations are as follows:

- The unique meaning of the status bits that reflects the PMU errors in data recording is not consistently implemented in various PMU brands and vintages. Defining the status bits with no possibility of misinterpretations and certifying the full compliance of a PMU to standard definitions by third-parties will be very helpful
- The event labels syntax and semantics are inconsistent across different recordings from various data contributors. Standardizing the event labeling is expected to tremendously improve the model selection and training
- The selection of times stamps for designating the event start and end time is made using a rather inaccurate timing source. Agreeing on which timing sources should be used would bring major improvement to the event characterization, which is a precursor for the successful implementation of data models

IV-b.3. Opportunities for further data model developments and improvements

The recommendations are as follows:

- Off-the-shelf machine learning models, while certainly a good starting point for education and training, are not going to achieve good performance for PMU data analytics without significant tuning
- The key challenges of AI/ML methods when it comes to analyzing power system data is in automating the data labeling, including time-stamping, as well as in capturing long data history
- The low-cost steps for utilities to take now to make the AI/ML approaches ready for big data analytics 2 or 3 years from now is to amend and open the field-recorded datasets for AI/ML experts to use for further studies.

APPENDIX A: Product or Technology Production

A-1. R. Baembitov, T. Dokic, M. Kezunovic and Z. Obradovic, “Fast Extraction and Characterization of Fundamental Frequency Events from a Large PMU Dataset Using Big Data Analytics,” Proc. 54th IEEE Hawaii International Conference on System Science (HICSS-54), Hawaii, USA, January 2021. pp. 3195-3204, doi: 10.24251/HICSS.2021.389

A-2. M. K. Alqudah, M. Pavlovski, T. Dokic, M. Kezunovic, Y. Hu and Z. Obradovic, "Fault Detection Utilizing Convolution Neural Network on Timeseries Synchrophasor Data From Phasor Measurement Units," in IEEE Transactions on Power Systems, doi: 10.1109/TPWRS.2021.3135336.

A-3. M. Pavlovski, M. Alqudah, T. Dokic, A. A. Hai, M. Kezunovic and Z. Obradovic, "Hierarchical Convolutional Neural Networks for Event Classification on PMU Measurements," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-13, 2021, Art no. 2514813, doi: 10.1109/TIM.2021.3115583.

A-4. A. Abdel Hai, T. Dokic, M. Pavlovski, M. Alqudah, M. Kezunovic, Z. Obradovic, “Transfer Learning for Event Detection from PMU Measurements with Scarce Labels,” IEEE Access, Vol. 9, 127420–127432, September 10 2021, doi: 10.1109/ACCESS.2021.3111727.

A-5. H. Otudi, T. Dokic, T. Mohamed, Y. Hu, M. Kezunovic, Z. Obradovic, “Line Faults Classification Using Machine Learning on Three Phases Voltages Extracted from Large Dataset of PMU Measurements,” Proc. 55th IEEE Hawaii International Conference on System Science (HICSS-55), Hawaii, USA, January 2022. DOI: 10.24251/HICSS.2022.425.

A-6. Z. Cheng, Y. Hu, Z. Obradovic, M. Kezunovic, “Using Synchrophasor Status Word as Data Quality Indicator: What to Expect in the Field?”, IEEE Smart Grid Synchronized Measurement and Analytics Conference, SGSMA 2022, Split, Croatia, May 2022. doi: 10.1109/SGSMA51733.2022.9806010

A-7. T. Dokic, et. all, “Machine Learning Using a Simple Feature for Detecting Multiple Types of Events From PMU Data,” IEEE Smart Grid Synchronized Measurement and Analytics Conference, SGSMA 2022, Split, Croatia, May 2022. doi: https://ieeexplore.ieee.org/abstract/document/9806000

A-8. M. Kezunovic, et al. “Automated System-wide Event Detection and Classification Using Machine Learning on Synchrophasor Data,” CIGRE General Session, Aug. 28-Sept 2, Paris, 2022.

A-9. M. Kezunovic, et. al. “Use of Machine Learning on PMU Data for Transmission System Fault Analysis,” CIGRE General Session, Aug. 28-Sept 2, Paris, 2022.

APPENDIX B: References

TR. Bandaragoda, KM. Ting, D. Albrecht, FT. Liu, Y. Zhu, JR. Wells, "Isolation-based anomaly detection using nearest-neighbor ensembles," *Computational Intelligence*, 2018; 34: 968– 998. https://doi.org/10.1111/coin.12156

M. Biswal, S. M. Brahma, and H. Cao, "Supervisory Protection and Automated Event Diagnosis Using PMU Data," *IEEE Transactions on Power Delivery*, vol. 31, no. 4, pp. 1855–1863, 2016.

M. Biswal, Y. Hao, P. Chen, S. Brahma, H. Cao, and P. De Leon, "Signal Features for Classification Of Power System Disturbances Using PMU Data," *Proc. 2016 Power Systems Computation Conference*, IEEE, 2016, pp. 1–7.

S. Brahma, et al., "Real-Time Identification of Dynamic Events in Power Systems Using PMU Data, and Potential Applications—Models, Promises, And Challenges," *IEEE Transactions on Power Delivery*, vol. 32, no. 1, pp. 294–301, 2016.

L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, 2001.

J. Browniee, "One-vs-Rest and One-vs-One for Multi-Class Classification," April 13, 2020. [Online].Available: https://machinelearningmastery.com/one-vs-rest-and-one-vs-one-for-multi-class-classification/

F. Ciancetta, et al., "A New Convolutional Neural Network-Based System for NILM Applications," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–12, 2021.

M. Cuturi, and M. Blondel. "Soft-DTW: A Differentiable Loss Function for Time-Series," *Proc. of the 34th Int. Conf. on Machine Learning* 70, pp. 894-903. JMLR. org, 2017.

O. P. Dahal, S. M. Brahma, and H. Cao, "Comprehensive Clustering of Disturbance Events Recorded by Phasor Measurement Units," *IEEE Transactions on Power Delivery*, vol. 29, no. 3, pp. 1390–1397, 2013.

O. P. Dahal et al., "Comprehensive Clustering of Disturbance Events Recorded by Phasor Measurement Units," *IEEE Trans. on Power Delivery*, vol. 29, no. 3, June 2014.

X. Ding and Q. He, "Energy-Fluctuated Multiscale Feature Learning with Deep Convnet For Intelligent Spindle Bearing Fault Diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 8, pp. 1926–1935, 2017.

S. Jafarzadeh; N. Moarref; Y. Yaslan; V. M. Istemihan Genc, "A CNN-Based Post-Contingency Transient Stability Prediction Using Transfer Learning," *Proc. 11th International Conference on Electrical and Electronics Engineering (ELECO)*, Bursa, Turkey, Nov. 2019.

M. Kesici, C. Saner, M. Mahdi, Y. Yaslan, and V. Genc. "Wide Area Measurement Based Online Monitoring and Event Detection Using Convolutional Neural Networks," *Proc. 2019 7th Int. Istanbul Smart Grids and Cities Congress and Fair (ICSG)*, pp. 223-227. IEEE, 2019

W.M. Kouw, M. Loog, "An Introduction To Domain Adaptation and Transfer Learning," arXiv:1812.11806, 2018.

W. Li, D. Deka, M. Chertkov, and M. Wang. "Real-Time Faulted Line Localization and PMU Placement in Power Systems Through Convolutional Neural Networks." *IEEE Trans. on Power Systems 34*, no. 6 (2019): 4640-4651.

W. Li, and M. Wang. "Identifying Overlapping Successive Events Using a Shallow Convolutional Neural Network," *IEEE Trans. on Power Systems 34*, no. 6 (2019): 4762-4772.

J. Ma et al., "Building Decision Trees for Characteristic Ellipsoid Method to Monitor Power System Transient Behaviors," *Proc. IEEE Power and Energy Society General Meeting*, Providence, RI, USA, July 2010.

J. C. Mayo-Maldonado, et al., "Data-Driven Framework to Model Identification, Event Detection, and Topology Change Location Using D-PMUs," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 9, pp. 6921–6933, 2020.

"Multiclass & Multilabel Classification with XGBoost [Online].Available:https://gabrielziegler3.medium. com/multiclass-multilabel-classification-with-xgboost

"Open-Source Gradient Boosting Library with Categorical Features Support," Yandex, [Online]. Available: https://catboost.ai/.

S.J. Pan, I.W. Tsang, J.T. Kwok, Q. Yang, "Domain Adaptation via Transfer Component Analysis," *IEEE Transactions on Neural Networks* 22(2):199–210, 2011.

L. Prokhorenkova et al., "CatBoost: Unbiased Boosting with Categorical Features," in *Advances in Neural Information Processing Systems* 31 (NIPS 2018), 2018.

Prudhvi, Poorna. "Support Vector Machines," Medium, Medium, 13 Oct. 2017, medium.com/@poornaprudhvi/support-vector-machines-bf0242ade522.

S. Ramaswamy, S. Rastogi, K. Shim, "Efficient Algorithms for Mining Outliers From Large Data Sets," *SIGMOD* Rec. 29, 2 (June 2000), 427–438, 2000. DOI: https://doi.org/10.1145/335191.335437

R. Sathya, A. Annamma, “Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification,” *International Journal of Advanced Research in Artificial Intelligence*, 2013, 2. 10.14569/IJARAI.2013.020206.

Scikit-learn – Machine Learning in Python, [Online] Available: https://scikit-learn.org/stable/.

Shahsavari, M. Farajollahi, E. M. Stewart, E. Cortez, and H. Mohsenian-Rad, “Situational Awareness in Distribution Grid Using Micro-PMU Data: A Machine Learning Approach,” *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 6167–6177, 2019.

J. Shi, B. Foggo, and N. Yu, “Power System Event Identification Based on Deep Neural Network with Information Loading” *IEEE Transactions on Power Systems*, vol. 14, no. 8, 2015.

“Sklearn.multiclass.OneVsOneClassifier,”Scikit,scikitlearn.org/stable/modules/generated/sklearn .multiclass. OneVsOneClassifier.html.

J. Van Haaren, A. Kolobov, J. Davis, “TODTLER: Two-Order-Deep Transfer Learning,” *Proc. 29th AAAI Conference on Artificial Intelligence*, 2015, 3007–3015.

V. Vercruyssen, W. Meert, G. Verbruggen, K. Maes, R. Bäumer, J. Davis, “Semi-Supervised Anomaly Detection with an Application to Water Analytics,” *Proc. IEEE International Conference on Data Mining (ICDM)*, Singapore, 2018, p527--536.

V. Vercruyssen, W. Meert, J. Davis, “Transfer Learning for Anomaly Detection through Localized and Unsupervised Instance Selection,” *Proc. AAAI Conference on Artificial Intelligence*, 2020.

S. Wang, P. Dehghanian, and L. Li. "Power Grid Online Surveillance through PMU-Embedded Convolutional Neural Networks," *IEEE Trans. on Industry Applications* 56, no. 2 (2019): 1146-1155.

W. Wang, H. Yin, C. Chen, A. Till, W. Yao, X. Deng, and Y. Liu. "Frequency Disturbance Event Detection Based on Synchrophasors and Deep Learning," *IEEE Trans. on Smart Grid* (2020).

L. Wen, X. Li, and L. Gao, “A New Two-Level Hierarchical Diagnosis Network Based on Convolutional Neural Network,” *IEEE Transactions on Instrumentation and Measurement,* vol. 69, no. 2, pp. 330–338, 2020.

XGBoost Model in Python.” *Machine Learning Mastery*, 18 Jan. 2021, machinelearningmastery.com/develop-first- xgboost-model-python-scikit-learn/.

T. Yiu, "Understanding Random Forest," *Medium, Towards Data Science*, June 12, 2019. [Online] towardsdatascience.com/understanding-random-forest-58381e0602d2.

Zakaria El Mrabet; Daisy Flora Selvaraj; Prakash Ranganathan, "Adaptive Hoeffding Tree with Transfer Learning for Streaming Synchrophasor Data Sets," *Proc. IEEE International Conference on Big Data (Big Data)*, Los Angeles, CA, USA, Dec. 2019

Y. Zhang, X. Wang, Y. Luo, Y. Xu, J. He, G. Wu, "A CNN Based Transfer Learning Method for High Impedance Fault Detection," *Proc. IEEE Power & Energy Society General Meeting (PESGM)*, Montreal, QC, Canada, Aug. 2020.

X. Zhu, A. Goldberg, "Synthesis Lectures on Artificial Intelligence and Machine Learning", 2009, Vol. 3, No. 1, Pages 1-130