From Big Data to Automated Event Characterization: How to Assure the Best Value from Synchrophasor Investments

Presenter: Mladen Kezunovic, PI
Texas A&M Engineering Experiment Station (TEES)

IEEE PES GM, Panel Session "Event Characterization Using Synchrophasor Big Data
Denver, Colorado, July

Overview

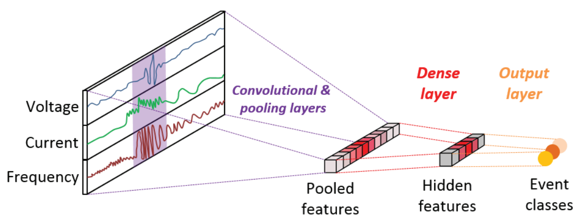
- Objectives:
  - Understand data properties
  - Set up data for model development and validation efforts
  - Develop models to detect, characterize and classify events
- Technical Approach:
  - Deploy the expert domain skills and ML/AI approaches to discover and utilize knowledge for development of data models
- Significance and Impact:
  - Develop useful models, offer recommendations for best practices, and facilitate initiation of new standardization efforts

Problem Statement

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*.

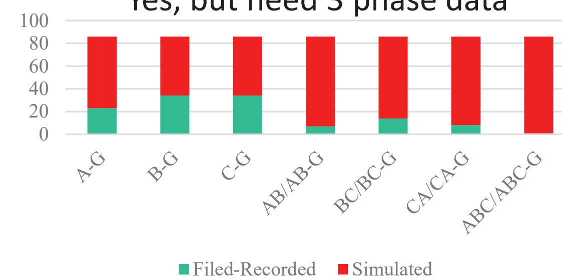
Q1: Can feature learning be automated?

Yes – Multi-channel filtering by CNN



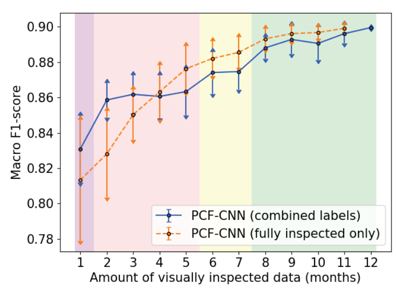
Q3: Can models be improved by using PMU data from simulations?

Yes, but need 3 phase data



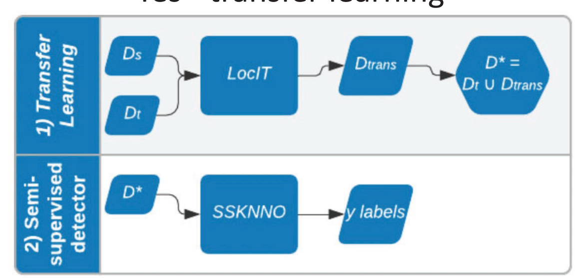
Q2: Should models learn from more data or from better data?

Use both if data is small

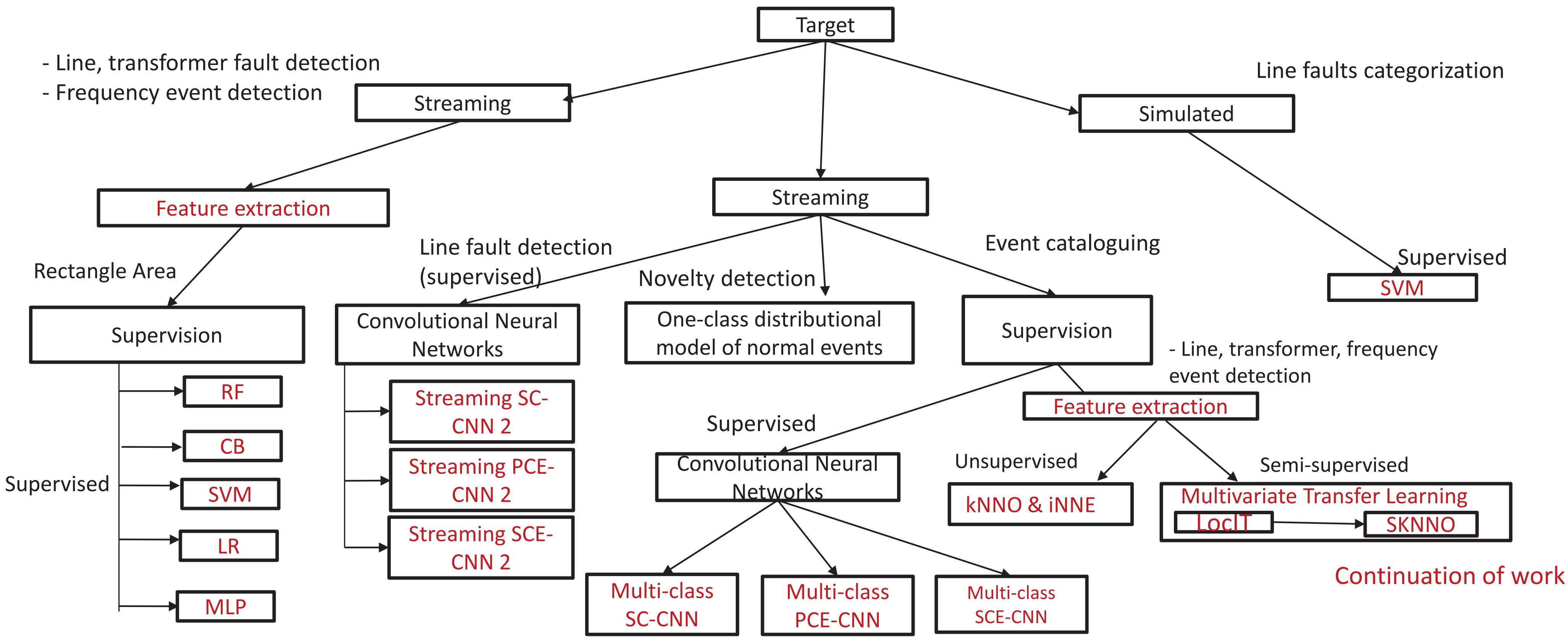


Q4: Can relevant labeled PMU data from a related task be used for learning on a new tasks?

Yes - transfer learning



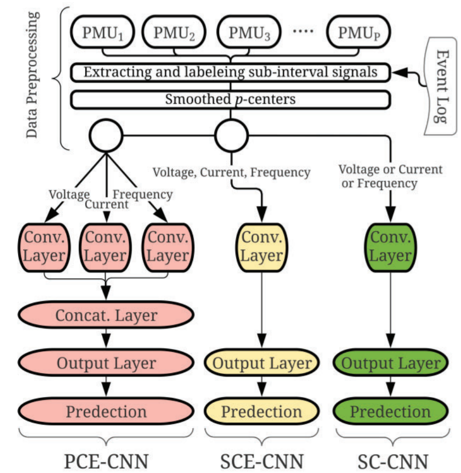
Experimental Approach



Convolution-based Fault Detection Utilizing Timeseries Synchrophasor Data from Phasor Measurement Units

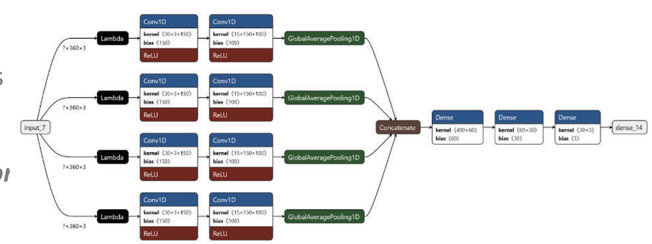
Q1: Can feature learning be automated? **Yes. Design automated event detection systems that doesn't rely on extensive manual study of data and feature engineering.**

- **Objective**: Develop an *end-to-end* supervised learning method with *automated feature learning* for fault detection (line faults) using sparse PMUs.
- **Challenges**: Sparse PMUs don't necessary cover all geo areas. Noisy data. Unknown event locations. Scalability of detection models.
- **Proposed solution:** Introduced a robust set of automated data preprocessing steps. Introduced CNN based models (PCE-CNN, SCE-CNN, SC-CNN), each introduces a different way of learning features.
- **Results**: Data from western interconnection were used. 2016 used for training and 2017 used for testing. Testing showed robust detection in multiple settings with AUC of 83%. More results are expected soon.



Extension: Voltage level aware CNN classification based on sliding window technique.

1. PMUs were split into voltage levels (134kV, 240kV, 300kV, 500kV) following a preprocessing scheme
2. From 55s Soft-DTW preprocessed signal, a 30s window is slid, taking those values as separate training instances for the 3 channels.
3. Sliding window data is passed to one of 3 CNN architectures (VL-CNN-MC is on the right). Each voltage level is processed separately before uniting the inferred features from all voltage. Results are in progress.



Publication: Alqudah, M., Pavlovski, M., Dokic, T., Kezunovic, M., Obradovic, Z. "***Convolution-based Fault Detection Utilizing Timeseries Synchrophasor Data from Phasor Measurement Units***," IEEE Transactions on Power Systems, 14 Dec, 2021 (Early Access)

QUANTA

Hierarchical Convolutional Neural Networks for Event Classification on PMU Measurements

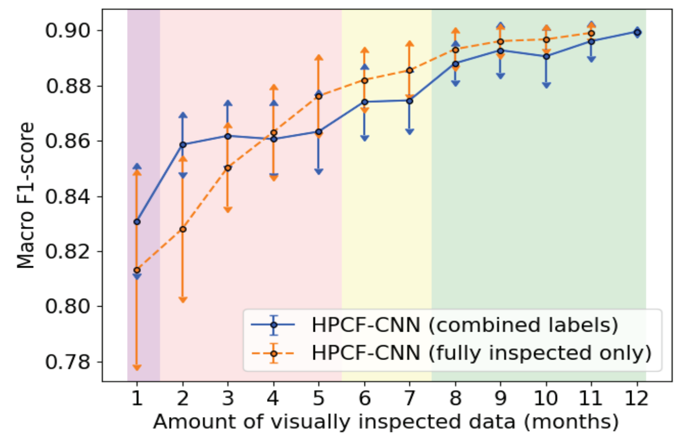
Q2: Should models learn from more data or from better data? Use both, if the data is small.

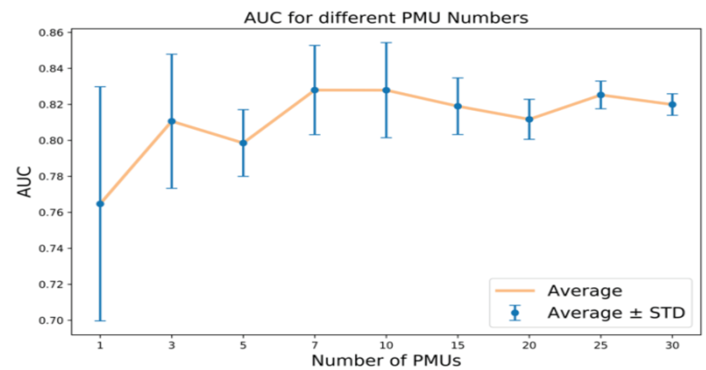
- **Objective:** Develop a multi-class event classification model using sparse PMUs. Where line-faults, fundamental frequency and normal events are detected.
- **Challenges:** Lack of high-quality labels, in-complete labeled data, inaccurate labels, noisy data, sparse PMUs.
- **Proposed Solution:** Utilize labels inspected by domain experts in addition to original labels. Introduce hierarchical convolutional neural networks.
- **Results:** Models were trained on 2016 data and tested on 2017 (Int. B). Accuracies of 89% to 94% were obtained depending on how much domain experts are involved. Manual labeling Scenarios:
  - **Scenario 1:** When domain expert's time is extremely limited

    Recommendation: Manually inspect and label at least 2 moths of data to receive satisfactory performance.
  - **Scenario 2:** When there are capacity for domain experts to fully inspect data

    Recommendation: Inspect ≥ of 8 months to receive high performance for automated detection models.

Publication: 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.





Line Faults Classification using Three Phases

Q3: ***Can models be improved by using PMU data from simulations*****? Yes.**

- **Objective**: Develop a supervised learning method to classify transmission line faults using PMU Measurement from two source of data.
- **Challenges 1:**It is difficult to separate PP (or 3P) from PP-G (or 3P-G) faults using field-recorded data.
- **Challenges 2:** There are fewer examples of PP, PPG,3P, and 3P-G faults compared to P-G faults in the field-recorded dataset.
- **Proposed solution.** Due to a limited number of examples per type of line faults in field-recorded PMU data, simulated data obtained from a synthetic system shown in Fig. 1 with much more prominent fault types are combined with field-recorded data to generate an integrated training set as shown in Fig 2.
- **Results.** Training the classification models with the combined dataset resulted in a classification accuracy of 98.58%. This is a significant improvement over 86.87% to 87.17% accuracy obtained by relying on the field-recorded dataset alone.

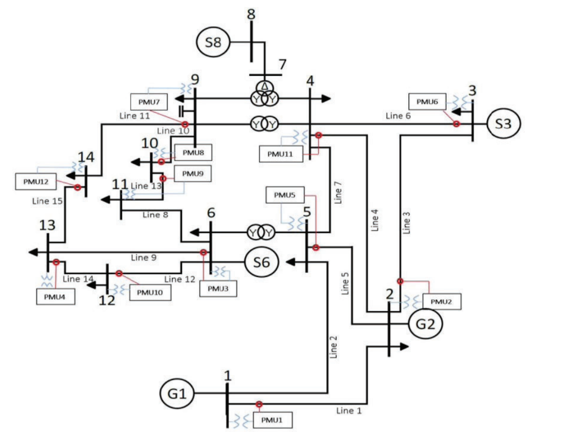


Figure 1:PMU placement in the synthetic IEEE 14-bus power system

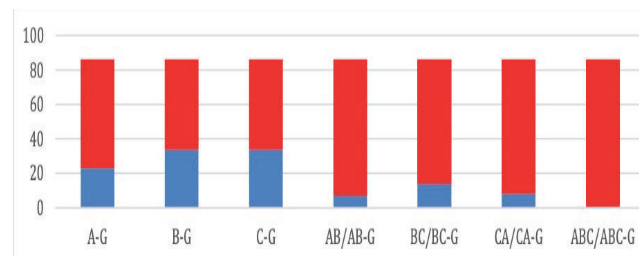


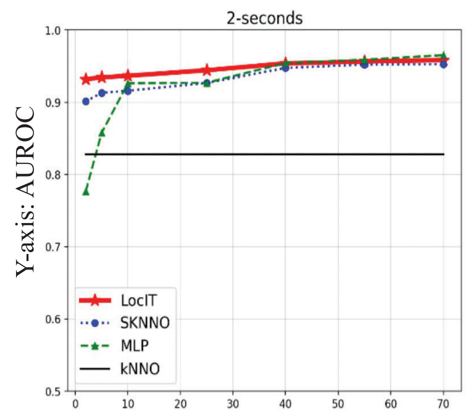
Figure 2: Integrated data distribution (training set)

Publication: 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 Int'l Conf. System Science*, Hawaii, USA, Jan. 2022.

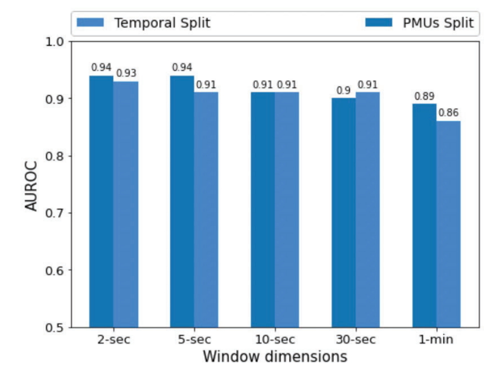
Transfer Learning for Event Detection from PMU Measurements with Scarce Labels

Q4: Can relevant labeled PMU data from a related task be used for learning on a new task? **Yes,** transfer learning.

- ***Objective: D***evelop a line fault, frequency, and transformer event detection method that is capable of detecting events based on minimal labeled data.
- **Challenges:** Supervised learning based event detection methods rely on adequate labeled data, which can be costly or infeasible to obtain. Leveraging data from a task to another might challenge supervised learning algorithms due to the concept and covariate shift assumptions.
- **Proposed solution:** Utilize transfer learning technique with semi-supervised learning detector to leverage a small number of labeled data from one task to the target domain without additional labeling effort, hence, detecting events without having to rely on event logs of PMU data.
- **Results:** Transfer learning method yields ~13% improvement in AUROC when compared to supervised learning algorithms. The performance is less affected by the decrease in the number of available labels.



X-axis: percentage of the labeled source data
LocIT: transfer learning; SKNNO: semi-supervised; MLP: supervised; kNNO: unsupervised



2%, 5%, 10%, 25%, 40%, 55%, and 70%, corresponding to 20, 51, 103, 259, 415, 570, and 726 of labeled source data

Feature Extraction:

$$RA_{PMU,TW} = (f_{max} - f_{min}) * (V_{max} - V_{min})$$

Publication: 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, 2021.

Outcomes

- Establishing facts:
  - Even now, let alone 5 years from now, collecting PMU data may have diminishing returns unless the analysis is automated
  - While automating the analysis of historical data has value, real benefit comes from predicting occurrence and mitigating impacts of undesirable events in real-time
- Offering Recommendations:
  - While utilities can gain by sharing data with each other, the real value is in sharing and following best practices in recording its own data, and then preserving it
- Facilitating standardization work:
  - Developing and adopting standardized approaches on PMU setting flags, the exact meaning of the error bits, and a common format for event labeling

Being Ready for ML & BD Analytics

- 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 tuning
- The key challenges of ML/AI 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
- When focusing on PMU data and event logs:
  - Improving data quality and recording practices could help in the development of ML/AI models in the future.
  - Data labeling should be done not only based on SCADA data but also based on data from other recording systems, including GPS time-stamps
  - It is essential to synergistically combine machine learning models with power systems domain knowledge since data-based models, as powerful as they are, will not be sufficient.
- The low-cost steps for utilities to take now to make the ML/AI approaches ready for big data analytics 2 or 3 years from now is to amend and open data sets for ML/AI experts to use.

Lessons Learned and Next Steps

- Improving data labeling, data quality and recording practices (three-phase vs positive sequence) is essential for future ML/AI applications in the utilities
- Providing power system topology and PMU placement helps in distinguishing power system events and assessing their importance/impact
- Using synthetic data has limited value except for the fault studies where the events are local, highly distinguishable, and resemble actual events closely.
- Data management of large data sets of streaming data is an expensive effort and requires new data management, data wrangling, and data viewing tools
- Enhancing PMU data with data from other utility recording devices/systems (DFRs, DPRs, SCADA), and with weather data can produce significant benefits
- Automating event detection and analysis, evolving from a posterior (historical) to a priori (predictive) formulation, requires further ML/AI research

Publications

Journals

- 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. DOI: 10.1109/TPWRS.2021.3135336.
- 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, pp. 1-13, no. 2514813, 2021.
- 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, September 2021.

Conferences

- M. Kezunovic, Z. Obradovic, Y. Hu “Automated System-wide Event Detection and Classification Using Machine Learning on Synchrophasor Data,” paper submitted to CIGRE General Session, Aug. 28-Sept 2, Paris, 2022.
- M. Kezunovic, Z. Obradovic, Y. Hu “Use of Machine Learning on PMU Data for Transmission System Fault Analysis,” paper submitted to CIGRE General Session, Aug. 28-Sept 2, Paris, 2022.
- 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
- T. Dokic, R. Baembitov, A. Abdel Hai, Z. Cheng, Y. Hu, M. Kezunovic, Z. Obradovic, “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
- H. Otudi, T. Dokic, T. Mohamed, M. Kezunovic, Y. Hu, Z. Obradovic, “Line Faults Classification Using Machine Learning on Three Phase Voltages Extracted from Large Dataset of PMU Measurements,” HICSS-55 Conference, Hawaii, USA, January 2022.
- 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,” HICSS-54 Conference, Hawaii, USA, January 2021.