Power System Feature-Based Event Classification by Means of Multiple PMU Data

1st Reza Nematirad
Quanta Technology, LLC
Raleigh, NC, USA
nematirad@ksu.edu

2nd Zheyuan Cheng
Quanta Technology, LLC
Raleigh, NC, USA
zcheng@quanta-technology.com

3rd Farrokh Aminifar
Quanta Technology, LLC
San Clemente, CA, USA
faminifar@quanta-technology.com

***Abstract*—Phasor Measurement Units (PMUs) provide time-synchronized measurements across the power grid, enabling data-driven event detection and classification for enhanced system monitoring and situational awareness. However, variations in event duration, spatial extent, and severity, along with coincident events, pose challenges for conventional classification models that require fixed-size inputs. This paper presents a feature-based framework that aggregates diverse attributes from all available PMUs for each event into a fixed-length vector, facilitating the application of standard machine learning classifiers, including Random Forest, XGBoost, and Multilayer Perceptron. A probabilistic post-processing scheme is further introduced to enable multi-label classification in the presence of overlapping events. Experiments using real-world PMU data demonstrate that the Random Forest model achieves 95% accuracy, while the proposed post-processing method yields an additional 3% improvement.**

***Index Terms*—Event classification, feature extraction, phasor measurement units**

I. Introduction

Phasor Measurement Units (PMUs) stream time-synchronized measurements—including voltage and current phasors, frequency, and rate of change of frequency (ROCOF)—from across the power grid to centralized systems. In addition to supporting a wide range of real-time protection, control, engineering, and planning applications, PMU data enable the development of data-driven methods for event detection and classification. Events occur frequently in power systems, and accurate detection and classification are essential for improving system situational awareness and reliability. While event detection has been extensively studied, event classification remains a challenging problem [1].

Each PMU produces a multivariate time series of dimension $[C \times T]$, where C denotes the number of measurement channels (e.g., voltage magnitude, phase angle) and T the number of time steps. Events are observed by varying numbers of PMUs; for instance, system-wide frequency disturbances are typically captured by many PMUs, whereas localized events such as faults or switching operations may only be detected by a few. Consequently, an event can be represented as $[P \times C \times T]$, where P is the number of PMUs involved. Moreover, event duration (T) varies with disturbance type. This high-dimensional structure, along with variations in the number of participating PMUs and event durations, presents significant challenges for conventional classification models that require uniformly sized input data.

Furthermore, multiple events may occur simultaneously, increasing the complexity of event classification. Conventional classifiers typically assign each instance to the class with the highest predicted probability [2]. However, overlapping events often yield comparable probabilities across multiple classes. For example, fault and voltage disturbance signatures may coexist within the same measurement. In such cases, a conventional classifier incorrectly assigns the instance to a single class—either fault or voltage disturbance—when it should, in fact, belong to both.

Several raw-data-based classification frameworks have been proposed for PMU event analysis. In [3], an autoencoder-based model is developed for event localization using simulated data from a modified IEEE 123-bus system, achieving 88.5% accuracy across 13 event types through a late-fusion strategy. A hierarchical convolutional neural network (CNN) is introduced in [4] to classify events directly from raw PMU time-series data. Evaluation on IEEE 118-bus simulated data demonstrates high accuracy across nine event categories. In [5], autoencoders combined with the Grey Wolf Optimizer are proposed to address class imbalance and data high-dimensionality. Ensemble models combining autoencoders and Gated Recurrent Units (GRUs) achieve strong performance on real PMU datasets. Similarly, [6] presents a cascaded residual 1D CNN integrated with fully connected layers, trained on simulated IEEE 39-bus PMU data to classify six power-system anomalies. Leveraging residual structures, the model attains 94.3% accuracy and outperforms conventional CNNs and shallow classifiers. Finally, [7] proposes a hierarchical CNN-based framework that eliminates explicit feature extraction and achieves high classification performance on RTDS-simulated IEEE 13-node data. However, these models often require uniformly sized input data.

To overcome the variable-input-size challenge, feature-based classification frameworks have been utilized that extract fixed-dimensional features to produce uniformly sized inputs. In [8], statistical and temporal features extracted from multiple PMUs are used for event classification employing Random

This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Solar Energy Technologies Office (SETO) Award Number DE-EE0011375. The views expressed herein do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

979-8-3315-8138-1

Forest and Support Vector Machine (SVM) models. The results demonstrate that the Random Forest classifier achieves higher accuracy and remains robust even with missing PMU data. A wavelet-based neural network approach utilizing a Probabilistic Neural Network (PNN) is presented in [9], achieving high classification accuracy across multiple power quality events using simulated PMU-like signals.

A Random Forest classifier is employed in [10] on features including envelope areas, statistical metrics, and oscillation characteristics, achieving 86% overall accuracy across multiple event categories on large-scale field data. In [11], an XGBoost-based classification framework is developed and trained on real PMU measurements from the IEEE 39-bus system to identify grid events such as regular operation, faults, line outages, generation changes, and load variations, demonstrating the superior performance of XGBoost. Similarly, [12] proposes a feature-based framework using Multilayer Perceptrons (MLPs) to classify fault, frequency, and oscillation events. While feature-based methods address the variable-input-size challenge through fixed-dimensional feature extraction, they inherently cannot account for overlapping events that coexist within the same measurement.

In this paper, we propose a feature-based classification approach that handles both variable input dimensions and overlapping events in PMU measurements. To address variability in data size, a comprehensive set of features is extracted, including basic statistical metrics, envelope areas over sliding windows, and diverse time- and frequency-domain characteristics. These features are computed across all PMUs for each event, concatenated into a fixed-length vector, and subsequently used to train machine learning models such as Random Forest, XGBoost, and MLP. To account for overlapping events, a probabilistic post-processing scheme is introduced to enable multi-label assignment. Instead of assigning a single label to each event, multiple labels are assigned when competing class probabilities are comparable. The key contributions of this study are summarized as follows:

1) A feature-based event classification framework is developed to manage high-dimensional PMU data with variable event durations and differing numbers of PMUs per event, addressing core challenges in real-world power system monitoring.
2) A probabilistic post-processing scheme is proposed to support multi-label classification for overlapping events.
3) Multiple machine learning models, including Random Forest, XGBoost, and MLP, are evaluated using real-world PMU datasets. Results demonstrate that the feature-based Random Forest classifier, combined with probabilistic post-processing, achieves the most robust and accurate performance.

II. Methodology

The following sections outline the data collection and methodology for classifying power system events by means of multiple PMU data.

A. Data Collection

In this study, a total of 210 real-world event records were collected from PMUs deployed across the U.S. Western Interconnection. The PMUs operate at a reporting rate of 60 samples per second and are part of a network comprising 35 units. Each event involves data from at least four PMUs, with some including measurements from all 35. For each PMU, six channels were utilized for event classification: three-phase voltage magnitudes (V_a, V_b, V_c) in per-unit, current positive-sequence magnitude (I_p), system frequency (f), and ROCOF. The temporal duration of events ranges from 1 to 27 minutes.

All events were manually verified and labeled by domain experts into five categories: voltage oscillation, frequency disturbance, voltage disturbance, fault, and transient events. Fig. 1 illustrates the per-unit phase-A voltage signals for representative events from each category. Each curve corresponds to measurements from an individual PMU. As shown, events differ in both spatial extent and temporal duration; for example, frequency disturbances are system-wide and observed by many PMUs, whereas transient events are localized and detected by only a few. The collected PMU data were subsequently used for feature extraction to support event classification.

B. Feature Extraction

To address the high-dimensional event structure and variability in both the number of PMUs per event and event duration, the raw PMU measurements were transformed into machine-learning-compatible representations through comprehensive feature extraction. A diverse set of features was

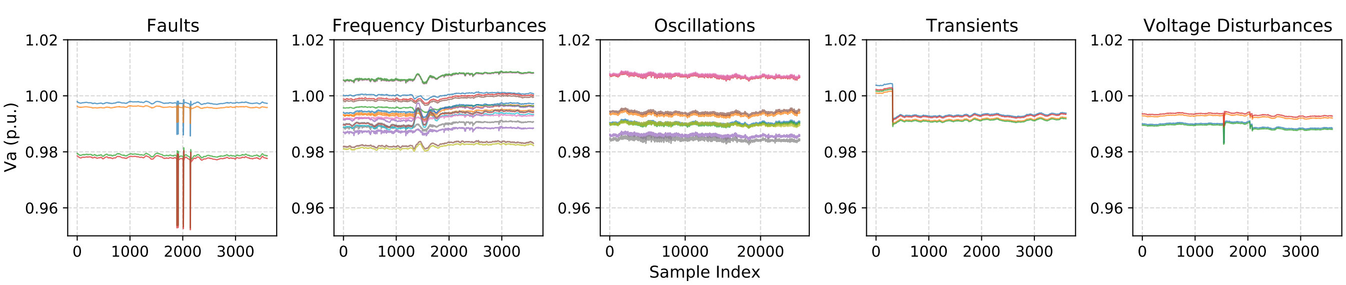

Fig. 1. Per-unit voltage magnitude of phase A from multiple PMUs for five event categories. Each curve represents measurements from one PMU. For clarity of illustration, other measurement channels are not shown.

979-8-3315-8138-1

derived to capture event dynamics, including basic statistical measures (mean, variance, standard deviation, and peak-to-peak value), envelope areas over sliding windows, and shape descriptors such as skewness, kurtosis, and impulse factor. Frequency-domain characteristics—including spectral centroid, flatness, entropy, flux, and dominant frequency—were obtained via the Fast Fourier Transform (FFT).

First- and second-order derivatives of voltage and frequency signals were computed to represent rate-of-change trends, while sag and swell durations, along with their corresponding magnitudes, were extracted to characterize transient behaviors during event transitions. The number of PMUs contributing to each event was also included as a contextual feature. All extracted features from individual PMUs were concatenated into a fixed-length vector, forming the input to machine learning classifiers.

C. Classification Algorithms

In this study, three machine learning algorithms including Random Forest, XGBoost, and MLP are employed to perform event classification.

1) Random Forest: Random Forest is an ensemble learning algorithm that constructs multiple decision trees using randomly sampled subsets of data and features. During classification, each tree independently predicts the event type based on the extracted features, and the final decision is obtained through majority voting across all trees. The method offers a transparent and interpretable framework for assessing feature importance [2]. Moreover, Random Forest is particularly well-suited for this study, as it demonstrates strong generalization capability and reduced susceptibility to overfitting when applied to moderate-sized or imbalanced datasets—conditions commonly encountered in real-world PMU event data.

2) Extreme Gradient Boosting: XGBoost is an optimized gradient boosting framework that constructs an ensemble of decision trees in a sequential manner, where each subsequent tree is trained to correct the residual errors of the preceding trees. Event classification is performed using the extracted feature vectors by optimizing a gradient-based objective function that minimizes classification loss. XGBoost effectively captures complex, nonlinear relationships among features through its additive tree structure and regularization mechanisms, often achieving high predictive accuracy and strong generalization performance [11].

3) Multilayer Perceptrons: MLPs are feedforward neural networks consisting of multiple layers of interconnected neurons. Each neuron computes a weighted sum of its inputs followed by a nonlinear activation function, enabling the network to model complex, nonlinear relationships among the extracted features. The MLP performs event classification by mapping input feature vectors to output event classes through successive hidden layers.

To evaluate the models' learning capability and generalization performance, learning curves are analyzed to compare training and validation behaviors across iterations.

D. Learning Curve Analysis

Learning curve analysis is employed to evaluate the generalization performance of the classification models. This analysis helps identify underfitting, where the model is too simple to capture the underlying data patterns, and overfitting, where the model learns noise or memorizes the training data instead of generalizing from it.

For the Random Forest and XGBoost models, learning curves are constructed by progressively increasing the training sample size and plotting classification accuracy on both the training and validation datasets. For the MLP, the learning curve tracks the training and validation loss across epochs. Ideally, a small and stable gap between the training and validation curves indicates good generalization, whereas large gaps or divergence suggest overfitting [2].

E. Evaluation Metrics

Model performance is evaluated using four standard metrics: **accuracy**, representing the ratio of correctly classified events to the total; **precision**, the proportion of correctly predicted samples for each class; **recall**, the ability to identify all true instances of a class; and the **F1-score**, the harmonic mean of precision and recall. After training the machine learning models, a probabilistic post-processing step is applied to evaluate and handle concurrent or overlapping events.

F. Probabilistic Post-Processing

Since multiple events may occur simultaneously, a probabilistic post-processing procedure is applied to enable multi-label assignment based on comparable predicted probabilities. For each data instance, if the predicted class probabilities exceed 30% and the difference between competing probabilities is less than 10%, the event is assigned to multiple classes. This dual-threshold strategy ensures that only sufficiently confident predictions (greater than 30%) are retained, while classes with similar likelihoods (difference less than 10%) are jointly recognized rather than arbitrarily separated. For example, if the classifier predicts an event as a fault with a probability of 42% and as a transient with a probability of 45%, both fault and transient labels are assigned to the event. This approach is particularly relevant in practical scenarios where a switching event often follows a fault within 100 milliseconds.

III. Results and Discussion

A. Dataset and Experimental Setup

The labeled dataset comprises 210 event samples, partitioned into 70% for training and 30% for testing. The training set is further stratified using 5-fold cross-validation, with each fold reserving 20% of the training data for validation. The dataset includes 11 fault events, 49 transients, 17 voltage disturbances, 8 frequency disturbances, and 125 oscillations.

Hyperparameter optimization is conducted using the Optuna framework, which efficiently explores the search space to identify optimal model configurations. The processed feature vectors extracted from all events are subsequently used to train the classification models [2].

979-8-3315-8138-1

TABLE I
CLASSIFICATION PERFORMANCE OF ALL MODELS ACROSS FIVE EVENT CLASSES.

Model	Acc.	Faults			Freq. Dist.			Oscillations			Transients			Volt. Dist.		
		Prec.	*Recall*	*F1*	*Prec.*	*Recall*	*F1*	*Prec.*	*Recall*	*F1*	*Prec.*	*Recall*	*F1*	*Prec.*	*Recall*	*F1*
Random Forest	0.95	0.83	1.00	0.91	1.00	1.00	1.00	0.95	1.00	0.97	0.93	0.93	0.93	1.00	1.00	1.00
XGBoost	0.82	1.00	1.00	1.00	1.00	1.00	1.00	0.83	0.92	0.87	0.82	0.64	0.72	0.06	0.50	0.11
MLP	0.75	0.67	1.00	0.80	0.00	0.00	0.00	0.80	0.98	0.88	0.80	0.62	0.70	0.00	0.00	0.00

B. Classification Results

The performance of all classification models is summarized in Table I. Among the evaluated models, the Random Forest achieved the highest overall accuracy of 95%, followed by XGBoost at 82% and the MLP at 75%. Random Forest demonstrated consistently strong performance across all event classes, achieving perfect recall and F1-scores for most categories. The optimal hyperparameters obtained using the Optuna framework are listed in Table II, enabling the Random Forest to maintain high accuracy and stable performance across all event types.

XGBoost also performed well in identifying faults, frequency disturbances, and oscillations, but showed weaker performance in detecting voltage disturbances and transient events. The MLP model exhibited difficulty in classifying frequency and voltage disturbance events, with both precision and recall equal to zero, indicating that none of the true samples or predictions for these classes were correctly identified. This behavior is likely due to the MLP's sensitivity to class imbalance in the training data, where these event types are underrepresented.

The learning curves of the machine learning models are presented in Fig. 2. For the Random Forest model, training accuracy decreases slightly while validation accuracy steadily improves as the training sample size increases, indicating enhanced generalization. The narrow gap between the two curves suggests balanced learning and minimal overfitting.

In contrast, XGBoost maintains a constant training accuracy of 1.0, while validation accuracy rises sharply from 0.44 to 0.83, suggesting potential overfitting caused by excessive memorization of the training data. For the MLP, the training loss continues to decrease throughout training, whereas the validation loss initially decreases and then increases slightly with additional epochs. This divergence indicates overfitting, as the model becomes increasingly specialized to the training data and less capable of generalizing to unseen samples.

TABLE II
OPTIMIZED HYPERPARAMETERS FOR THE RANDOM FOREST CLASSIFIER

Parameter	Value	Parameter	Value
Number of trees	137	Bootstrap sampling	True
Maximum tree depth	8	Splitting criterion	gini
Min. samples to split	5	Class weight	balanced
Min. samples per leaf	2	Maximum features	log2

C. Feature Importance Analysis

The ten most important features identified by the Random Forest model are presented in Fig. 3. The results show that the most influential features are primarily derived from frequency- and voltage-related measurements. Among them, the ROCOF spectral flux ranks highest, indicating that rapid variations in the spectral content of ROCOF are strong discriminators of dynamic events. Voltage-based features, including voltage variance and the first- and second-order voltage derivatives, also contribute significantly. These features capture the magnitude and rate of voltage fluctuations, making them particularly effective for detecting faults and voltage disturbances characterized by abrupt amplitude changes. Overall, the feature importance analysis demonstrates that the Random Forest model relies on physically interpretable indicators that align with known power system dynamics.

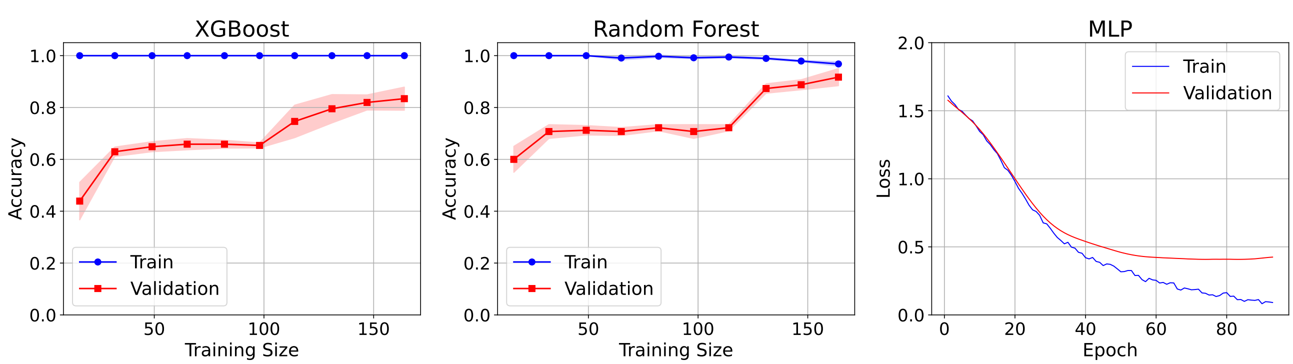

Fig. 2. Learning curves of Random Forest, XGBoost, and MLP models showing training and validation performance trends.

979-8-3315-8138-1

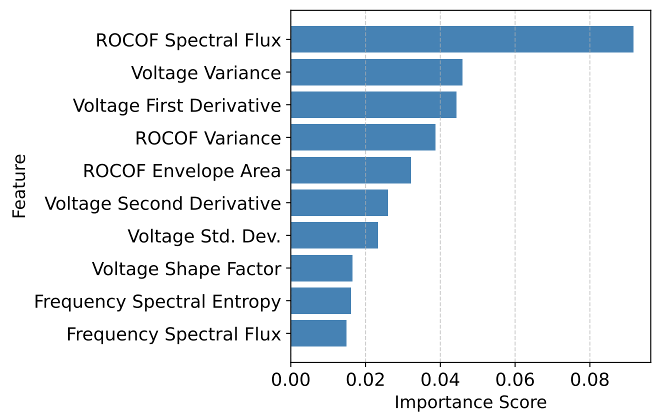

Fig. 3. Top ten most important features identified by the Random Forest.

D. Probabilistic Post-Processing Analysis

As discussed earlier, certain events may overlap and exhibit characteristics corresponding to multiple classes. For instance, in the Random Forest results, one event labeled as an oscillation was misclassified as a transient, with the model assigning probabilities of 43.9% for oscillation and 44.99% for transient. Fig. 4 illustrates the corresponding phase-A voltage and ROCOF signals for this event, which physically exhibits attributes of both categories. Another event in the dataset satisfies similar probabilistic conditions. When both events are reassigned to multiple classes using the proposed post-processing scheme, the Random Forest model's accuracy increases from 95% to 98%. One remaining misclassification does not meet these probabilistic thresholds and is likely attributable to labeling inconsistencies in the dataset.

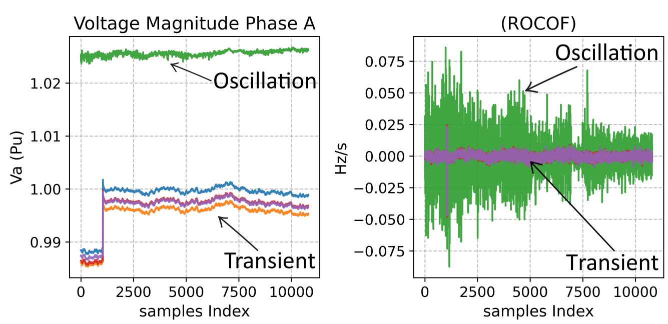

Fig. 4. Voltage phase A and ROCOF of the oscillation–transient overlapping event.

IV. Conclusion

This study presented a classification framework designed to address key challenges in real-world PMU event analysis, including variability in the number of PMUs per event, differences in event duration, and the occurrence of overlapping events. To mitigate data size variability, a comprehensive set of features was extracted from all available PMUs for each event and concatenated into a fixed-length representation for model training. A probabilistic post-processing scheme was introduced to enable multi-label assignment when class probabilities were comparable.

Experimental evaluation on real-world PMU data demonstrated that the feature-based Random Forest classifier achieved the highest accuracy and interpretability, attaining 95% overall accuracy and outperforming both XGBoost and MLP models. The proposed probabilistic post-processing method further enhanced the model's ability to identify overlapping events, improving classification performance.

Future research may explore normalization and dimensionality reduction techniques to further address data variability, as well as time-series classification models capable of directly processing raw, variable-length PMU sequences as an alternative to feature-based approaches.

References

[1] T. Mohamed, M. Kezunovic, Z. Obradovic, Y. Hu, and Z. Cheng, "Application of machine learning to oscillation detection using pmu data based on prony analysis," in *2022 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*. IEEE, 2022, pp. 1–5.

[2] R. Nematirad, M. Behrang, and A. Pahwa, "Acoustic-based online monitoring of cooling fan malfunction in air-forced transformers using learning techniques," *IEEE Access*, vol. 12, pp. 26 384–26 400, 2024.

[3] I. Niazazari and H. Livani, "A pmu-data-driven disruptive event classification in distribution systems," *Electric Power Systems Research*, vol. 157, pp. 251–260, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S037877961730500X

[4] M. Pavlovski, M. Alqudah, T. Dokic, A. A. Hai, M. Kezunovic, and Z. Obradovic, "Hierarchical convolutional neural networks for event classification on pmu measurements," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–13, 2021.

[5] D. Lal and R. Varadarajan, "Optimised autoencoder-based ensemble deep learning approaches for cyber-physical event classification utilizing synchrophasor pmu data," *Results in Engineering*, p. 105884, 2025.

[6] X. Lei, H. Ma, B. Liu, and Z. Li, "Power system events classification technology based on deep-learning," in *2024 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2024, pp. 1–5.

[7] M. N. I. Siddique, M. Shafiullah, S. Mekhilef, H. Pota, and M. Abido, "Fault classification and location of a pmu-equipped active distribution network using deep convolution neural network (cnn)," *Electric Power Systems Research*, vol. 229, p. 110178, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378779624000671

[8] Y. Liu, L. Yang, A. Ghasemkhani, H. Livani, V. A. Centeno, P.-Y. Chen, and J. Zhang, "Robust event classification using imperfect real-world pmu data," *IEEE Internet of Things Journal*, vol. 10, no. 9, pp. 7429–7438, 2023.

[9] Z.-L. Gaing, "Wavelet-based neural network for power disturbance recognition and classification," *IEEE Transactions on Power Delivery*, vol. 19, no. 4, pp. 1560–1568, 2004.

[10] Z. CHENG, J. J. Schmidt, Y. Hu, D. Hart, and J. Holbach, "Automated high-speed power system event detection and classification using synchrophasor data," Oct. 10 2024, US Patent App. 18/131,155.

[11] M. S. Abdalzaher, M. Shaaban, and R. Aburukba, "Using machine learning for classifying pmus observation in smart grids," in *2024 25th International Middle East Power System Conference (MEPCON)*. IEEE, 2024, pp. 1–6.

[12] T. Mohamed, M. Kezunovic, Y. Paudel, V. Vittal, A. Pal, D. Joshi, K. Menuka, M. Venkatasubramanian, and G. Torresan, "Automated detection and classification of critical power system events using ml on pmu data," in *2024 International Conference on Smart Grid Synchronized Measurements and Analytics (SGSMA)*. IEEE, 2024, pp. 1–6.

979-8-3315-8138-1